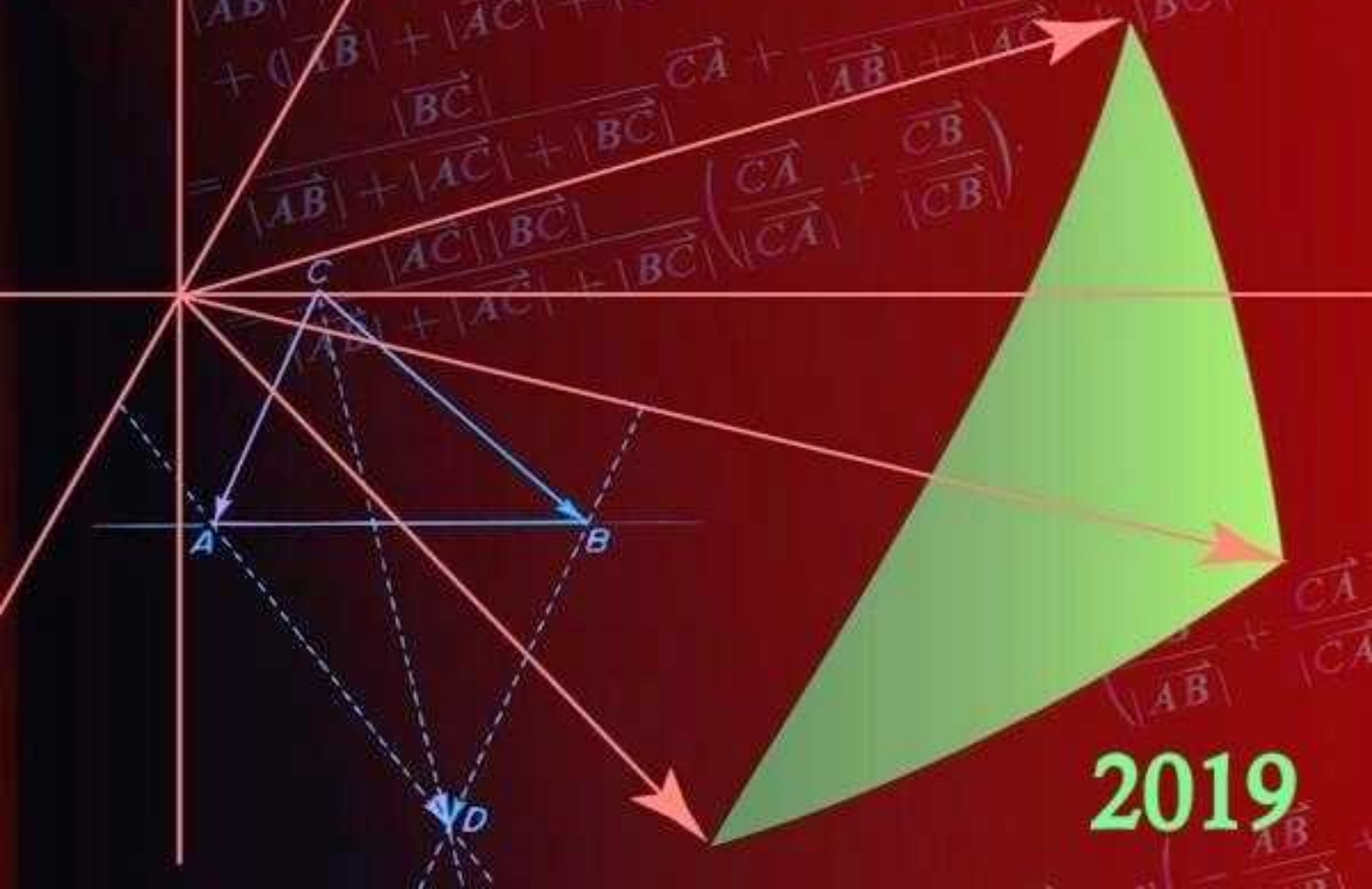
MATHEMATICS

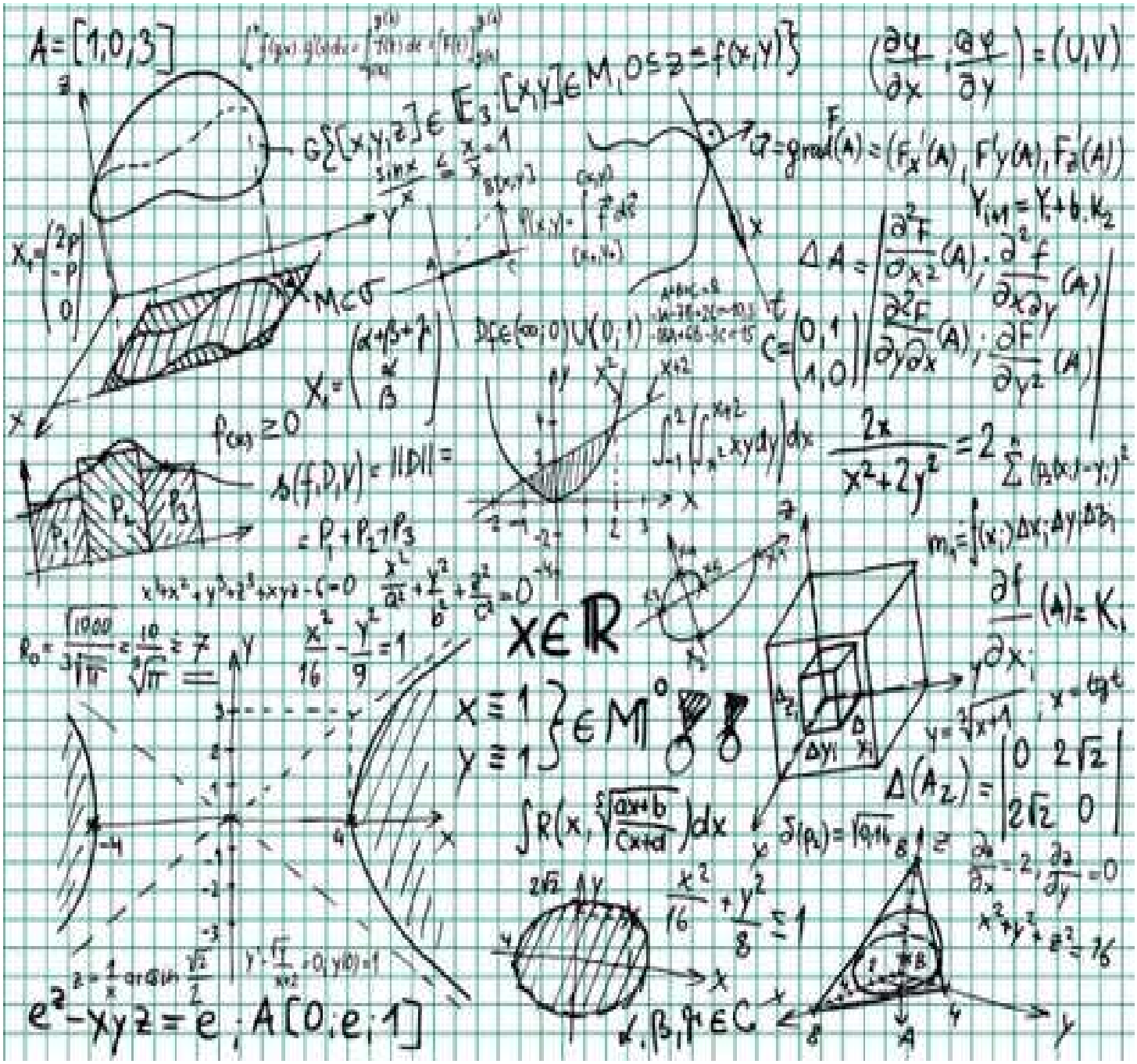
VECTORS

and Their Applications

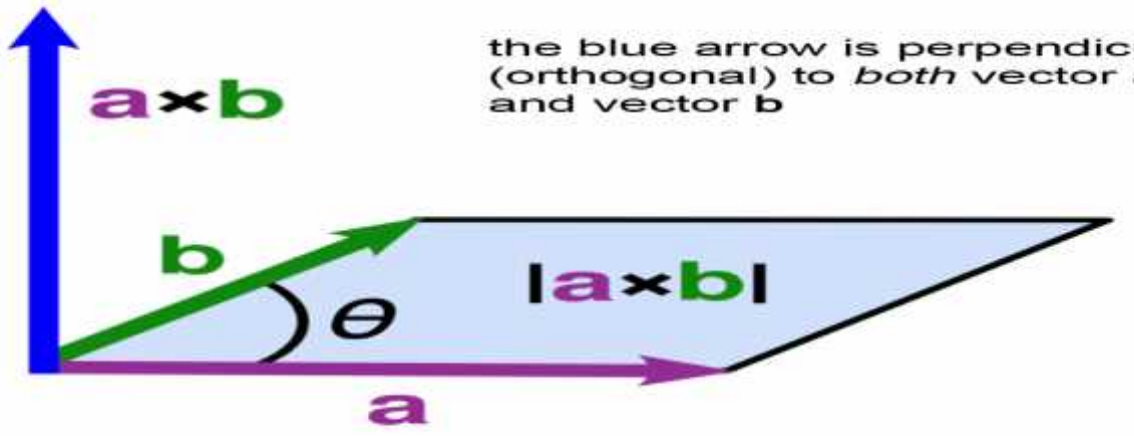
Nouh Hassan



2019



blue arrow is the resultant vector, with scalar value a times b times sine(θ), which is the area of the parallelogram in the plane of **a** and **b**



Cross-product of two vectors a and b separated by angle θ. Vertical vector is the cross-product value & direction

All rights are reserved

Translated by:

Diab Publishing & Distribution

www.Diab-publishing.com

Preface

This book contains valuable lessons of scientific value in mathematics, especially in relation to the wide field of sports (vectors).

This book is the unique encyclopedia of mathematics that helps in the teaching profession and in the culture of mathematics and in the contemporary aspects of this book with the various sciences of physics, engineering and technical sciences.

The researcher Mr. Nouh Hassan has relied on special curricula for educational education, research education and mathematics. It is a riches for the pioneers of mathematics and the world of the comprehensive scientific level.

Author Nouh Hassan

1/9/2019

Mr. Nouh Hassan in brief:

- Born in 1957, Maarake Village, Tyre in southern Lebanon.
- He received his primary and intermediate education at the official battle school and finished high school at the official secondary school.
- He graduated from the Teachers 'and Teachers' College in Nabatieh in 1979.
- He obtained his educational leave from the Lebanese University in mathematics in 1985.
- He was ranked first in Mathematics in the Lebanese Civil Service Council in 1995.
- He completed his graduation from the Faculty of Education at the Lebanese University in 1996 with the rank of secondary professor in the staff of the Ministry of National Education.
- Participated in many educational activities, training courses and public events.

Contents

Chapters **Pages**

1. Introduction to vectors……..…1

2. Vectors and Applications…….……..…25

3. Vectors and Scalars…….……...…53

4. Vectors and Translation Maps.....82

5. Vectors Algebra…….…....105

6. Vectors in Euclidean space……..….240

7. Vectors and Linear Equations…...……..……..304

8. Theory of Space of the Vectors…….....438

Chapter One: Introduction to vectors

A vector is a quantity that has both a magnitude (or size) and a direction. Both of these properties must be given in order to specify a vector completely. In this unit we describe how to write down vectors, how to add and subtract them, and how to use them in geometry.

In order to master the techniques explained here it is vital that you undertake plenty of practice exercises so that they become second nature.

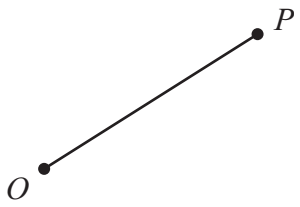
After reading this text, and/or viewing the video tutorial on this topic, you should be able to:

- distinguish between a vector and a scalar;
- understand how to add and subtract vectors;
- know when one vector is a multiple of another;
- use vectors to solve simple problems in geometry.

1. Introduction

Vector quantities are extremely useful in physics. The important characteristic of a vector quantity is that it has both a magnitude (or size) and a direction. Both of these properties must be given in order to specify a vector completely.

An example of a vector quantity is a displacement. This tell us how far away we are from a fixed point, and it also tells us our direction relative to that point.



Another example of a vector quantity is velocity. This is speed, in a particular direction. An example of velocity might be 60 mph due north.

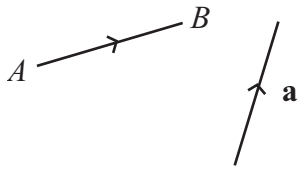
A quantity with magnitude alone, but no direction, is not a vector. It is called a *scalar* instead. One example of a scalar is distance. This tells us how far we are from a fixed point, but does not give us any information about the direction. Another example of a scalar quantity is the mass of an object.

Key Point

A vector has both magnitude and direction, and both these properties must be given in order to specify it. A quantity with magnitude but no direction is called a scalar.

2. Representing vector quantities

We can represent a vector by a line segment. This diagram shows two vectors.

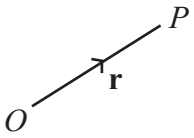


We have used a small arrow to indicate that the first vector is pointing from A to B. A vector pointing from B to A would be going in the opposite direction.

Sometimes we represent a vector with a small letter such as $\mathbf{a}$, in a bold typeface. This is common in textbooks, but it is inconvenient in handwriting. In writing, we normally put a bar underneath, or sometimes on top of, the letter: $\underline{a}$ or $\overline{a}$. In speech, we call this the vector "a-bar".

3. Position vectors

Sometimes vectors are referred to a fixed point, an origin. Such a vector is called a position vector. So we might refer to the position vector of a point P with respect to an origin O. In writing, might put $\overline{OP}$ for this vector. Alternatively, we could write it as $\mathbf{r}$. These two expressions refer to the same vector.



4. Some notation for vectors

What does it mean if, for two vectors, $\mathbf{a} = \mathbf{b}$? This means first that the length of $\mathbf{a}$ equals the length of $\mathbf{b}$, so that the two vectors have the same magnitude. But it also means that $\mathbf{a}$ and $\mathbf{b}$ are in the same direction. How can we write this down more succinctly?

If two vectors are "in the same direction", then they are parallel. We write this down as $\mathbf{a}//\mathbf{b}$.

For length, if we have a vector $\overline{AB}$, we can write its length as AB without the bar. Alternatively, we can write it as $|\overline{AB}|$. The two vertical lines give us the modulus, or size of, the vector. If we have a vector written as $\mathbf{a}$, we can write its length as either $|\mathbf{a}|$ with two vertical lines, or as a in ordinary type (or without the bar). This is why it is very important to keep to the convention that has been adopted in order to distinguish between a vector and its length.

Key Point

The length of a vector $\overline{AB}$ is written as

$$AB \text{ or } |\overline{AB}|,$$

and the length of a vector $\mathbf{a}$ is written as

$$a \text{ (in ordinary type, or without the bar) or as } |\mathbf{a}|.$$

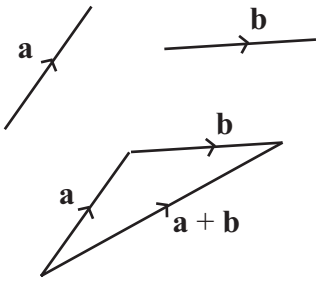
If two vectors $\mathbf{a}$ and $\mathbf{b}$ are parallel, we write

$$\mathbf{a}//\mathbf{b}$$

5. Adding two vectors

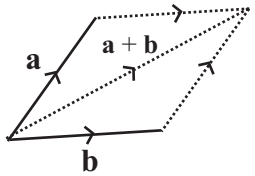
One of the things we can do with vectors is to add them together. We shall start by adding two vectors together. Once we have done that, we can add any number of vectors together by adding the first two, then adding the result to the third, and so on.

In order to add two vectors, we think of them as displacements. We carry out the first displacement, and then the second. So the second displacement must start where the first one finishes.



The sum of the vectors, $\mathbf{a} + \mathbf{b}$ (or the *resultant*, as it is sometimes called) is what we get when we join up the triangle. This is called the *triangle law* for adding vectors.

There is another way of adding two vectors. Instead of making the second vector start where the first one finishes, we make them both start at the same place, and complete a parallelogram. This is called the *parallelogram law* for adding vectors. It gives the same result as the triangle law, because one of the properties of a parallelogram is that opposite sides are equal and in the same direction, so that $\mathbf{b}$ is repeated at the top of the parallelogram.

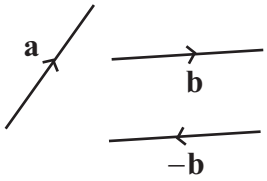


Key Point

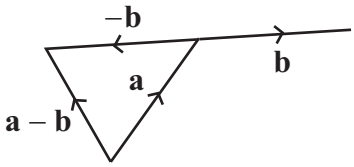
We can add two vectors $\mathbf{a}$ and $\mathbf{b}$ by making $\mathbf{b}$ start where $\mathbf{a}$ finishes, and completing the triangle. Alternatively, we can make $\mathbf{a}$ and $\mathbf{b}$ start at the same place, and take the diagonal of the parallelogram.

6. Subtracting two vectors

What is $\mathbf{a}-\mathbf{b}$? We think of this as $\mathbf{a}+(-\mathbf{b})$, and then we ask what $-\mathbf{b}$ might mean. This will be a vector equal in magnitude to $\mathbf{b}$, but in the reverse direction.



Now we can subtract two vectors. Subtracting $\mathbf{b}$ from $\mathbf{a}$ will be the same as *adding* $-\mathbf{b}$ to $\mathbf{a}$.

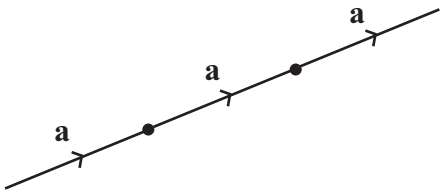


Key Point

$\mathbf{a}-\mathbf{b}$ means $\mathbf{a}+(-\mathbf{b})$

7. Adding a vector to itself

What happens when you add a vector to itself, perhaps several times? We write, for example, $\mathbf{a}+\mathbf{a}+\mathbf{a}=3\mathbf{a}$.



In the same way, we would write

$$n\mathbf{a}=\underbrace{\mathbf{a}+\ldots+\mathbf{a}}_{n \text{ copies}}.$$

Key Point

A vector $n\mathbf{a}$ is in the same direction as the vector $\mathbf{a}$, but n times as long.

8. Vectors of unit length

There is one more piece of notation we shall use when writing vectors. If $\mathbf{a}$ is any vector, we shall write $\hat{\mathbf{a}}$ to represent a unit vector in the direction of $\mathbf{a}$. A unt vector is a vector whose length is 1, so that

$$|\hat{\mathbf{a}}| = 1.$$

This notation gives us another way of writing the vector $\mathbf{a}$: we can write it as $a\,\hat{\mathbf{a}}$, so that it is the length a multiplied by the unit vector $\hat{\mathbf{a}}$.

Key Point

A unit vector in the direction of the vector $\mathbf{a}$ is written as $\hat{\mathbf{a}}$, so that

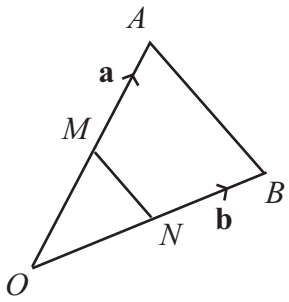
$$\mathbf{a} = a\,\hat{\mathbf{a}}.$$

9. Using vectors in geometry

Example

There is a useful theorem in geometry called the *mid-point theorem*. In this theorem, we take two points A and B, defined with respect to an origin O. Let us write $\mathbf{a}$ for the position vector of A, and $\mathbf{b}$ for the position vector of B. We can join A and B with a line, to give a triangle.

Now take the mid-point M of the line OA, and the mid-point N of the line OB, and join M to N with a line. Can we say anything about the relationship between the line MN and the line AB?



We can answer this very easily with vectors. We can write the vector for the line segment $\overline{AB}$ as $\overline{AO} + \overline{OB}$. Now $\overline{AO}$ is the reverse of the vector $\mathbf{a}$, so it is $-\mathbf{a}$. And $\overline{OB}$ is the same as the vector $\mathbf{b}$. Therefore

$$\begin{aligned} \overline{AB} &= \overline{AO} + \overline{OB} \\ &= (-\mathbf{a}) + \mathbf{b} \\ &= \mathbf{b} - \mathbf{a}. \end{aligned}$$

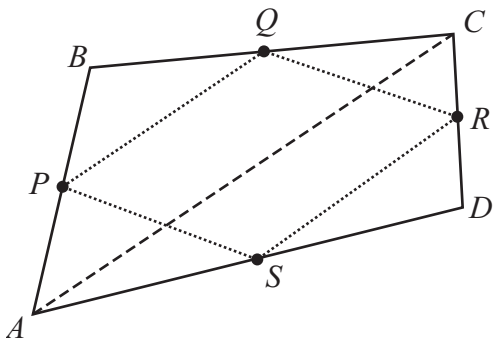
What about $\overline{MN}$? Following the same reasoning, this is $\overline{MO} + \overline{ON}$. But what is $\overline{MO}$? This is a vector half the length of $\overline{AO}$, and in the same direction, so it must be $\frac{1}{2}(-\mathbf{a})$. In the same way, $\overline{ON}$ is in the same direction as $\overline{OB}$, but is half the length, so it must be $\frac{1}{2}\mathbf{b}$. Therefore

$$\begin{aligned} \overline{MN} &= \overline{MO} + \overline{ON} \\ &= \tfrac{1}{2}(-\mathbf{a}) + \tfrac{1}{2}\mathbf{b} \\ &= \tfrac{1}{2}(\mathbf{b} - \mathbf{a}). \end{aligned}$$

Now we can compare $\overline{AB}$ and $\overline{MN}$. From our calculation, we can see that $\overline{MN}$ is $\frac{1}{2}\overline{AB}$. So, as this is a vector equation, it tells us two things. First, it tells us about magnitude, so that $MN = \frac{1}{2}AB$. Also, it tells us that MN and AB must be in the same direction, so that $MN // AB$. This is called the mid-point theorem for a triangle. It states that if you join the mid-points of two sides of a triangle then the resulting line is equal to half of the third side of the triangle, and is parallel to it.

Example

We can apply the mid-point theorem to a quadrilateral, or indeed to any four points in space, to give an interesting geometrical result. We shall call the four points A, B, C and D. We shall also give labels to the mid-points of the four sides: we shall call the mid-points P, Q, R and S. Now let us join the four mid-points, to make a new shape $PQRS$. What kind of shape is this?



We can identify the shape by joining the points A and C.

If we apply the mid-point theorem to triangle ABC, we see that

$$\overline{PQ} = \tfrac{1}{2}\overline{AC}.$$

But if we apply the mid-point theorem to the triangle ADC, we also see that

$$\overline{RS} = \tfrac{1}{2}\overline{AC}.$$

If we combine these two equations, we then obtain

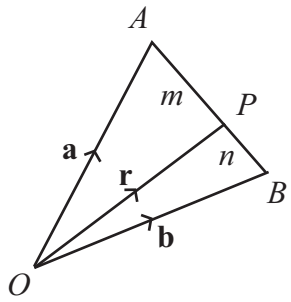
$$\overline{PQ} = \overline{RS}.$$

Now this is a vector equation, and so it tells us two things. First, it tells us that the length of PQ is the same as the length of RS. And secondly, it tells us that the direction of PQ is the same as the direction of RS, so that PQ and RS are parallel. But having two parallel sides of equal length is a property which defines a parallelogram, and so the shape $PQRS$ must be a parallelogram.

Example

We shall now use vectors to prove one more theorem.

Take two points A and B, having position vectors $\mathbf{a}$, $\mathbf{b}$ with respect to an origin O. Draw the line AB, and take a point P on that line which divides it in the ratio of m to n. What is the position vector of P with respect to O?.



We can use the same method that we used before. We know that

$$\overline{OP} = \overline{OA} + \overline{AP}, \tag{1}$$

and we also know that $\overline{OA} = \mathbf{a}$. But what is $\overline{AP}$?

Now $\overline{AP}$ is in the same direction as $\overline{AB}$, and their lengths are in the ratio of m to $m+n$. So

$$\overline{AP} = \frac{m}{m+n}\,\overline{AB}. \tag{2}$$

We also know that

$$\begin{aligned} \overline{AB} &= \overline{AO} + \overline{OB} \\ &= \mathbf{b} - \mathbf{a}. \end{aligned}$$

Now we can put these three statements together, replacing $\overline{AP}$ in equation (1) by using equation (2), and replacing $\overline{AB}$ in equation (2) by using the equation (3), so that everything will be written in terms of $\mathbf{a}$ and $\mathbf{b}$. This gives us

$$\overline{OP} = \mathbf{a} + \frac{m}{m+n}(\mathbf{b} - \mathbf{a}).$$

Putting all this over a common denominator then gives

$$\overline{OP} = \frac{(m+n)\mathbf{a} + m(\mathbf{b} - \mathbf{a})}{m+n}.$$

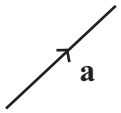
If we expand the brackets, the term $m\mathbf{a}$ will cancel with the term $m(-\mathbf{a})$, and so finally we have

$$\overline{OP} = \frac{n\mathbf{a} + m\mathbf{b}}{m+n}.$$

This formula gives us a way of calculating the position vector of the point P. For instance, if m and n were both 1 then P would be the mid-point of AB. The position vector of the midpoint would be $(\mathbf{a} + \mathbf{b})/2$. As another example, if $m = 2$ and $n = 1$, so that P was two-thirds of the way along the line, then the position vector of P would be $(\mathbf{a} + 2\mathbf{b})/3$.

Exercises

1. The vector $\mathbf{a}$ is shown below.

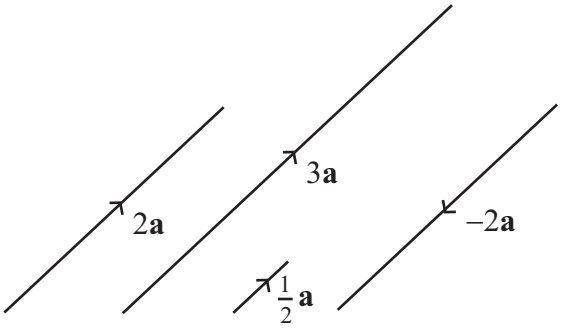


Sketch the vectors $2\mathbf{a}$, $3\mathbf{a}$, $\frac{1}{2}\mathbf{a}$ and $-2\mathbf{a}$.

2. In ΔOAB, $\overline{OA} = \mathbf{a}$ and $\overline{OB} = \mathbf{b}$. In terms of $\mathbf{a}$ and $\mathbf{b}$,
   (a) What is $\overline{AB}$?
   (b) What is $\overline{BA}$?
   (c) What is $\overline{OP}$, where P is the midpoint of AB?
   (d) What is $\overline{AP}$?
   (e) What is $\overline{BP}$?
   (f) What is $\overline{OQ}$, where Q divides AB in the ratio 2:3?
3. What is meant by a unit vector?
4. If $\mathbf{e}$ is a unit vector, what is the length of $3\mathbf{e}$?
5. In ΔABC, $AB = \mathbf{a}$, $BC = \mathbf{b}$, $CA = \mathbf{c}$. What is $\mathbf{a} + \mathbf{b} + \mathbf{c}$?

Answers

1.



2.
(a) $\mathbf{b}-\mathbf{a}$ (b) $\mathbf{a}-\mathbf{b}$ (c) $\frac{1}{2}(\mathbf{a}+\mathbf{b})$ (d) $\frac{1}{2}(\mathbf{b}-\mathbf{a})$
(e) $\frac{1}{2}(\mathbf{a}-\mathbf{b})$ (f) $\frac{3}{5}\mathbf{a}+\frac{2}{5}\mathbf{b}$

3. A vector with length 1

4. 3

5. 0

1. Introduction

It is useful to separate physical quantities into two types: the first are called **scalars**; the second are known as **vectors**. A scalar is a quantity that can be described by a single number which can be positive, negative or zero. An example of a scalar quantity is the mass of an object, so we might state that 'the mass of the stone is 3 kg'. It is important to give the units in which the quantity is measured. Obvious examples of scalars are temperature and length, but there are many other engineering applications in which scalars play an important role. For example, speed, work, voltage and energy are all scalars.

On the other hand, vectors are quantities which require the specification of a **magnitude** and a **direction**. An example of a vector quantity is the force applied to an object to make it move. When the object shown in Figure 1 is moved by applying a force to it we achieve different effects depending on the direction of the force.

Figure 1: A force is a vector quantity

In order to specify the force completely we must state not only its magnitude (its 'strength') but also the direction in which the force acts. For example we might state that 'a force of 5 newtons is applied vertically from above'. Clearly this force would have a different effect from one applied horizontally. The *direction* in which the force acts is crucial.

There are many engineering applications where vectors are important. Force, acceleration, velocity, electric and magnetic fields are all described by vectors. Furthermore, when computer software is written to control the position of a robot, the position is described by vectors.

Sometimes confusion can arise because words used in general conversation have specific technical meanings when used in engineering calculations. An example is the use of the words 'speed' and 'velocity'. In everyday conversation these words have the same meaning and are used interchangeably. However in more precise language they are not the same. **Speed** is a scalar quantity described by giving a single number in appropriate units; for example 'the speed of the car is 40 kilometres per hour'. On the other hand **velocity** is a vector quantity and must be specified by giving a direction as well. For example 'the velocity of the aircraft is 20 metres per second due north'.

In engineering calculations, the words speed and velocity cannot be used interchangeably. Similar problems arise from use of the words 'mass' and 'weight'. **Mass** is a scalar which describes the amount of substance in an object. The unit of mass is the kilogramme. **Weight** is a vector, the direction of which is vertically downwards because weight arises from the action of gravity. The unit of weight is the newton. **Displacement** and **distance** are related quantities which can also cause confusion. Whereas distance is a scalar, displacement is 'directed distance', that is, distance together with a specified direction. So, referring to Figure 2, if an object is moved from point A to point B, we can state that the distance moved is 10 metres, but the displacement is 10 metres *in the direction from A to B*.

You will meet many other quantities in the course of your studies and it will be helpful to know which are vectors and which are scalars. Some common quantities and their type are listed in Table 1. The S.I. units in which these are measured are also shown.

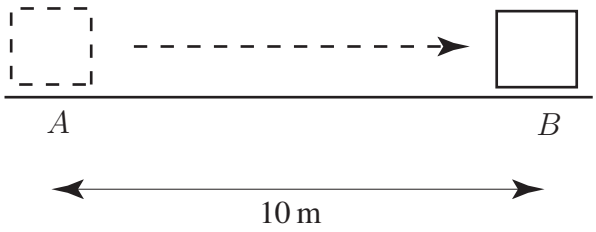


Figure 2: Displacement means directed distance

Table 1. Some common scalar and vector quantities

quantity	type	S.I. unit
distance	scalar	metre, m
mass	scalar	kilogramme, kg
temperature	scalar	kelvin, K
pressure	scalar	pascal, Pa
work	scalar	joule, J
energy	scalar	joule, J
displacement	vector	metre m
force	vector	newton, N
velocity	vector	metres per second, m s^{-1}
acceleration	vector	metres per second per second, m s^{-2}

Exercise

State which of the following are scalars and which are vectors:

(a) the volume of a petrol tank,
(b) a length measured in metres,
(c) a length measured in miles,
(d) the angular velocity of a flywheel,
(e) the relative velocity of two aircraft,
(f) the work done by a force,
(g) electrostatic potential,
(h) the momentum of an atomic particle.

Answer

(a), (b), (c) (f), (g) are scalars.

(d), (e), and (h) are vectors

2. The mathematical description of vector quantities

Because a vector has a direction as well as a magnitude we can represent a vector by drawing a line. The length of the line represents the **magnitude** of the vector given some appropriate scale, and the direction of the line represents the **direction** of the vector. We call this representation a **directed line segment**. For example, Figure 3 shows a vector which represents a velocity of 3 m s^{-1} north-west. Note that the arrow on the vector indicates the direction required.

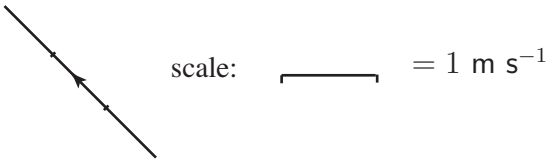


Figure 3: A vector quantity can be represented by drawing a line

More generally, Figure 4 shows an arbitrary vector quantity.

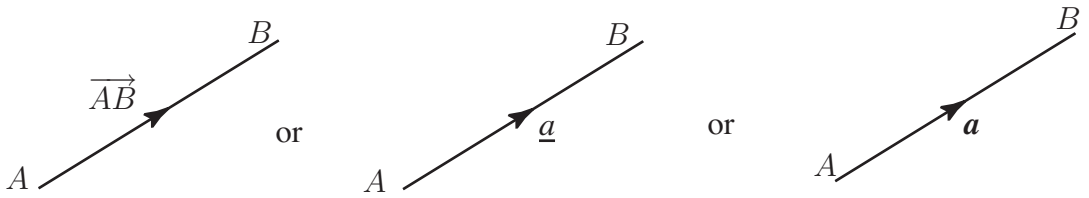


Figure 4: Representation of vectors

It is important when writing vectors to distinguish them from scalars. Various notations are used. In Figure 4 we emphasise that we are dealing with the vector from A to B by using an **arrow** and writing $\overrightarrow{AB}$. Often, in textbooks, vectors are indicated by using a **bold typeface** such as $\boldsymbol{a}$. It is difficult when handwriting to reproduce the bold face and so it is conventional to **underline** vector quantities and write $\underline{a}$ instead. So $\overrightarrow{AB}$ and $\underline{a}$ represent the same vector in Figure 4. We can also use the notation $\underline{AB}$. In general in this Workbook we will use underlining but we will also use the arrow notation where it is particularly helpful.

Example 1

Figure 5 shows an object pulled by a force of 10 N at an angle of 60° to the horizontal.

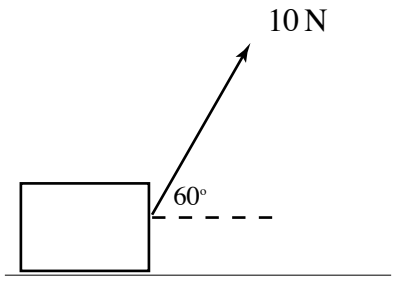


Figure 5

Show how this force can be represented by a vector.

Solution

The force can be represented by drawing a line of length 10 units at an angle of $60°$ to the horizontal, as shown below.

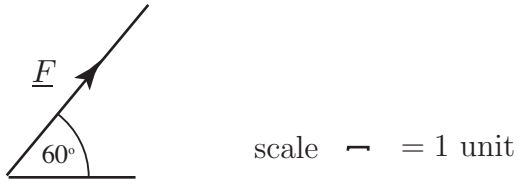


Figure 6

We have labelled the force $\underline{F}$. When several forces are involved they can be labelled $\underline{F}_1$, $\underline{F}_2$ and so on.

When we wish to refer simply to the magnitude (or length) of a vector we write this using the **modulus** sign as $|\overrightarrow{AB}|$, or $|\underline{a}|$, or simply a (without the underline.)

In general two vectors are said to be **equal vectors** if they have the same magnitude and same direction. So, in Figure 7 the vectors $\overrightarrow{CD}$ and $\overrightarrow{AB}$ are equal even though their locations differ.

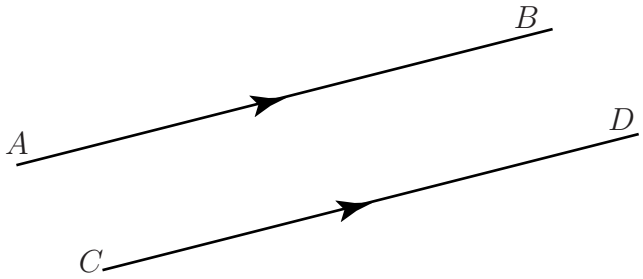


Figure 7: Equal vectors

This is a useful and important property of vectors: a vector is defined only by its direction and magnitude, not by its location in space. These vectors are often called **free** vectors.
The vector $-\underline{a}$ is a vector in the opposite direction to $\underline{a}$, but has the same magnitude as $\underline{a}$, as shown in Figure 8.

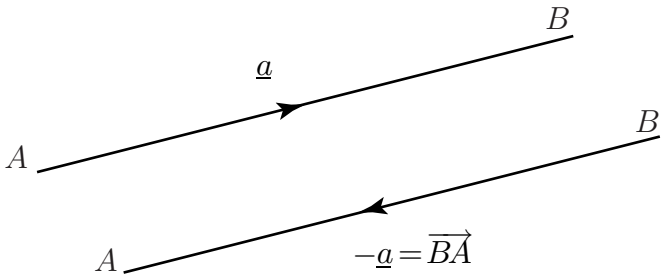


Figure 8

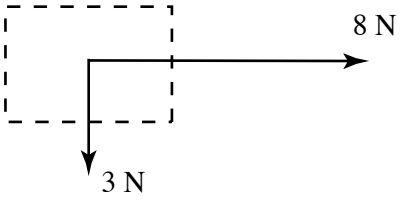
Geometrically, if $\underline{a} = \overrightarrow{AB}$ then $-\underline{a} = \overrightarrow{BA}$.

Exercises

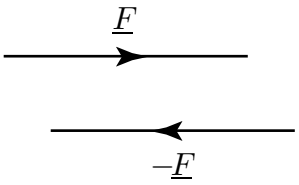
1. An object is subject to two forces, one of 3 N vertically downwards, and one of 8 N, horizontally to the right. Draw a diagram representing these two forces as vectors.
2. Draw a diagram showing an arbitrary vector $\underline{F}$. On the diagram show the vector $-\underline{F}$.
3. Vectors $\underline{p}$ and $\underline{q}$ are equal vectors. Draw a diagram to represent $\underline{p}$ and $\underline{q}$.
4. If $\underline{F}$ is a vector, what is meant by F?

Answers

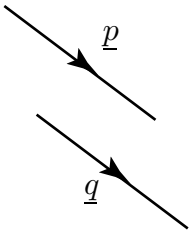
1.



2.



3.



4. F is the magnitude of $\underline{F}$.

3. Addition of vectors

Vectors are added in a particular way known as the **triangle law**. To see why this law is appropriate to add them this way consider the following example:

Example: The route taken by an automated vehicle

An unmanned vehicle moves on tracks around a factory floor carrying components from the store at A to workers at C as shown in Figure 9.

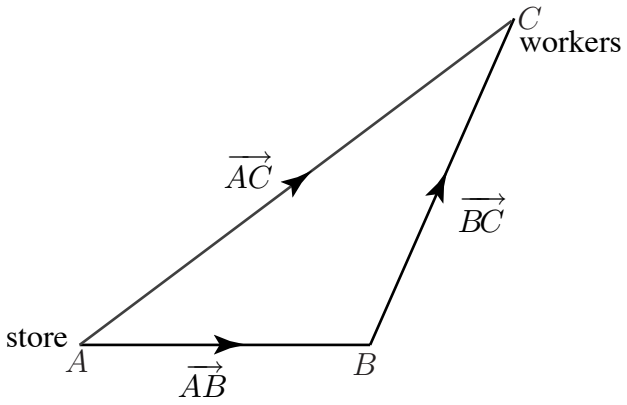


Figure 9

The vehicle may arrive at C either directly or via an intermediate point B. The movement from A to B can be represented by a displacement vector $\overrightarrow{AB}$. Similarly, movement from B to C can be

represented by the displacement vector $\overrightarrow{BC}$, and movement from A to C can be represented by $\overrightarrow{AC}$. Since travelling from A to B and then B to C is equivalent to travelling directly from A to C we write

$$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$$

This is an example of the triangle law for adding vectors. We add vectors $\overrightarrow{AB}$ and $\overrightarrow{BC}$ by placing the tail of $\overrightarrow{BC}$ at the head of $\overrightarrow{AB}$ and completing the third side of the triangle so formed ($\overrightarrow{AC}$).

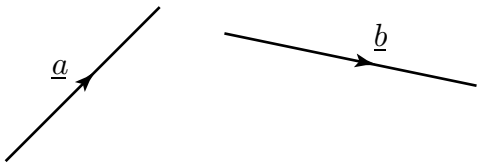


Figure 10: Two vectors $\underline{a}$ and $\underline{b}$

Consider the more general situation in Figure 10. Suppose we wish to add $\underline{b}$ to $\underline{a}$. To do this $\underline{b}$ is translated, keeping its direction and length unchanged, until its tail coincides with the head of $\underline{a}$. Then the sum $\underline{a} + \underline{b}$ is defined by the vector represented by the third side of the completed triangle, that is $\underline{c}$ in Figure 11. Note, from Figure 11, that we can write $\underline{c} = \underline{a} + \underline{b}$ since going along $\underline{a}$ and then along $\underline{b}$ is equivalent to going along $\underline{c}$.

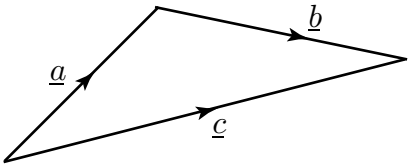
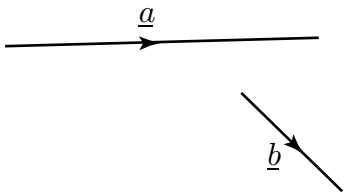


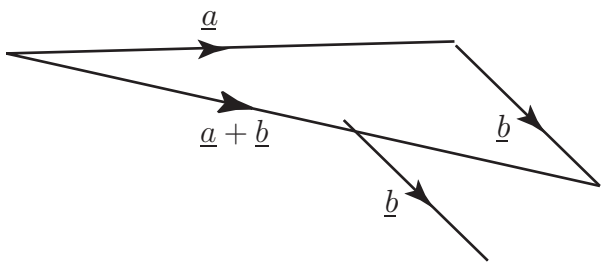
Figure 11: Addition of the two vectors of Figure 10 using the triangle law

Using vectors $\underline{a}$ and $\underline{b}$ shown below, draw a diagram to find $\underline{a} + \underline{b}$. Find also $\underline{b} + \underline{a}$. Is $\underline{a} + \underline{b}$ the same as $\underline{b} + \underline{a}$?



Your solution

Answer



yes $\underline{a}+\underline{b}=\underline{b}+\underline{a}$

It is possible, using the triangle law, to prove the following rules which apply to any three vectors $\underline{a}$, $\underline{b}$ and $\underline{c}$:

Key Point 1

Vector Addition

$\underline{a}+\underline{b}=\underline{b}+\underline{a}$	vector addition is commutative
$\underline{a}+(\underline{b}+\underline{c})=(\underline{a}+\underline{b})+\underline{c}$	vector addition is associative

Example: Resultant of two forces acting upon a body

A force $\underline{F}_1$ of 2 N acts vertically downwards, and a force $\underline{F}_2$ of 3 N acts horizontally to the right, upon the body shown in Figure 12.

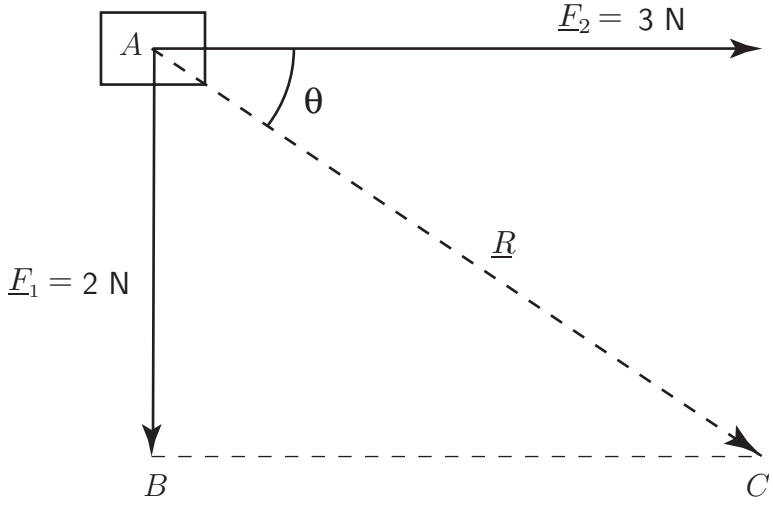


Figure 12

We can use vector addition to find the combined effect or **resultant** of the two concurrent forces. (Concurrent means that the forces act through the same point.) Translating $\underline{F}_2$ until its tail touches

the head of $\underline{F_1}$, we complete the triangle ABC as shown. The vector represented by the third side is the resultant, $\underline{R}$. We write

$$\underline{R} = \underline{F_1} + \underline{F_2}$$

and say that $\underline{R}$ is the **vector sum** of $\underline{F_2}$ and $\underline{F_1}$. The resultant force acts at an angle of θ below the horizontal where $\tan\theta = 2/3$, so that $\theta = 33.7°$, and has magnitude (given by Pythagoras' theorem) $\sqrt{13}$ N.

Example: Resolving a force into two perpendicular directions

In the previous Example we saw that two forces acting upon a body can be replaced by a single force which has the same effect. It is sometimes useful to reverse this process and consider a single force as equivalent to two forces acting at right-angles to each other.

Consider the force $\underline{F}$ inclined at an angle θ to the horizontal as shown in Figure 13.

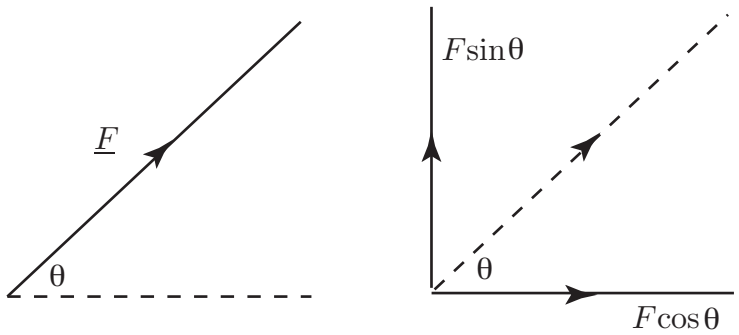


Figure 13

$\underline{F}$ can be replaced by two forces, one of magnitude $F\cos\theta$ and one of magnitude $F\sin\theta$ as shown. We say that $\underline{F}$ has been **resolved into two perpendicular components**. This is sensible because if we re-combine the two perpendicular forces of magnitudes $F\cos\theta$ and $F\sin\theta$ using the triangle law we find $\underline{F}$ to be the resultant force.

For example, Figure 14 shows a force of 5 N acting at an angle of $30°$ to the x axis. It can be resolved into two components, one directed along the x axis with magnitude $5\cos 30°$ and one perpendicular to this of magnitude $5\sin 30°$. Together, these two components have the same effect as the original force.

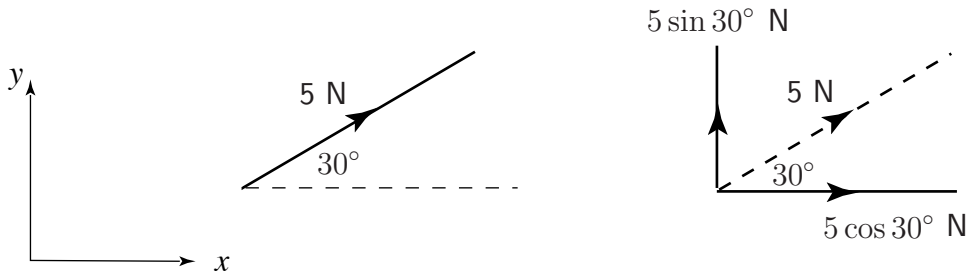
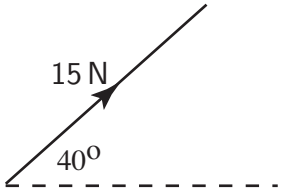


Figure 14

Consider the force shown in the diagram below.



Resolve this force into two perpendicular components, one horizontally to the right, and one vertically upwards.

Your solution

Answer

Horizontal component is $15\cos 40^\circ$ N $= 11.50$ N; vertical component is $15\sin 40^\circ$ N $= 9.64$ N

The need to **resolve** a vector along a given direction occurs in other areas. For example, as a police car or ambulance with siren operating passes by the pitch of the siren appears to increase as the vehicle approaches and decrease as it goes away. This change in pitch is known as the **Doppler effect** This effect occurs in any situation where waves are reflected from a moving object.

A radar gun produces a signal which is bounced off the target moving vehicle so that when it returns to the gun, which also acts as a receiver, it has changed pitch. The speed of the vehicle can be calculated from the change in pitch. The speed indicated on the radar gun is the speed directly towards or away from the gun. However it is not usual to place oneself directly in front of moving vehicle when using the radar gun (Figure 15(a).) Consequently the gun is used at an angle to the line of traffic (Figure 15(b).)

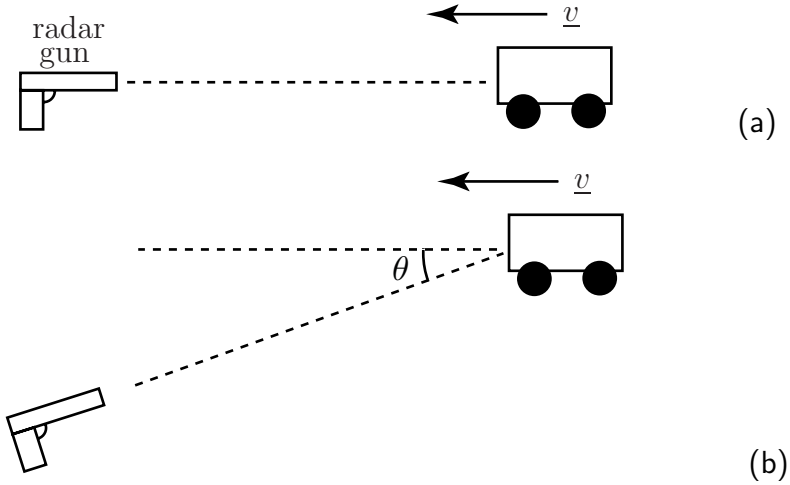


Figure 15

This means that it registers only the component of the velocity towards the gun. Suppose that the true speed along the road is v. Then the component measured by the gun ($v\cos\theta$) is less than v.

Example 2

A safety inspector wishes to check the speed of a train along a straight piece of track. She stands 10 m to the side of the track and uses a radar gun. If the reading on the gun is to be within 5% of the true speed of the train, how far away from the approaching train should the reading be taken?

Solution

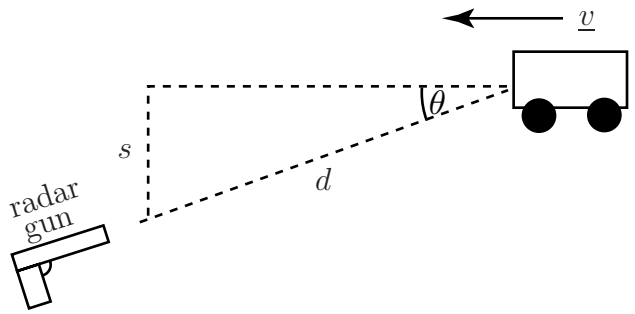


Figure 16

For an error of 5%, the gun should read $0.95|\underline{v}|$. So

$$|\underline{v}|\cos\theta = 0.95|\underline{v}| \quad \text{or} \quad \cos\theta = 0.95$$

If the distance to the side of the track at which the gun is used is s m and the distance between the radar gun and the train is d m, then

$$\sin\theta = \frac{s}{d}$$

Here $s = 10$ so

$$\sqrt{1-\cos^2\theta} = \frac{10}{d}.$$

So, if $\cos\theta = 0.95$

$$d = \frac{10}{\sqrt{1-(0.95)^2}} = 32.03$$

This means that the reading should be taken when the train is over 32 m from the radar gun to ensure an error of less than 5%.

The force vectors on an aeroplane in steady flight

The forces acting on an aeroplane are shown in Figure 17.

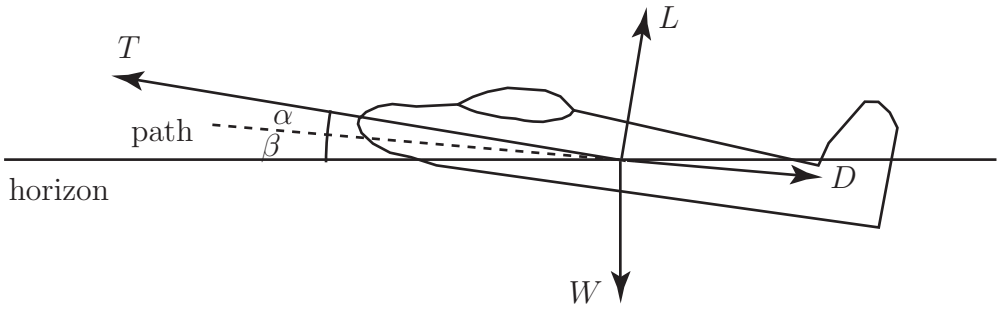


Figure 17

The **magnitude** (strength) of the forces are indicated by

T: the thrust provided by the engines,

W: the weight,

D: the drag (acting against the direction of flight) and

L: the lift (taken perpendicular to the path.)

In a more realistic situation force vectors in *three* dimensions would need to be considered. These are introduced later in this Workbook.

As the plane is in *steady* flight the sum of the forces in any direction is zero. (If this were not the case, then, by Newton's second law, the non-zero resultant force would cause the aeroplane to accelerate.)

So, resolving forces in the direction of the path:

$$T\cos\alpha - D - W\sin\beta = 0$$

Then, resolving forces perpendicular to the path:

$$T\sin\alpha + L - W\cos\beta = 0$$

If the plane has mass 72 000 tonnes, the drag is 130 kN, the lift is 690 kN and $\beta = 6^o$ find the magnitude of the thrust and the value of α to maintain steady flight. From these two equations we see:

$$T\cos\alpha = D + W\sin\beta = 130000 + (72000)(9.81)\sin 6^o = 203830.54$$

and

$$T\sin\alpha = W\cos\beta - L = (72000)(9.81)\cos 6^o - 690000 = 12450.71$$

hence

$$\tan\alpha = \frac{T\sin\alpha}{T\cos\alpha} = \frac{12450.71}{203830.54} = 0.061084 \quad \rightarrow \quad \alpha = 3.50^0$$

and consequently, for the thrust:

$T = 204210$ N.

4. Subtraction of vectors

Subtraction of one vector from another is performed by adding the corresponding negative vector. That is, if we seek $\underline{a} - \underline{b}$ we form $\underline{a} + (-\underline{b})$. This is shown geometrically in Figure 18. Note that in the right-hand diagram the arrow on $\underline{b}$ has been reversed to give $-\underline{b}$.

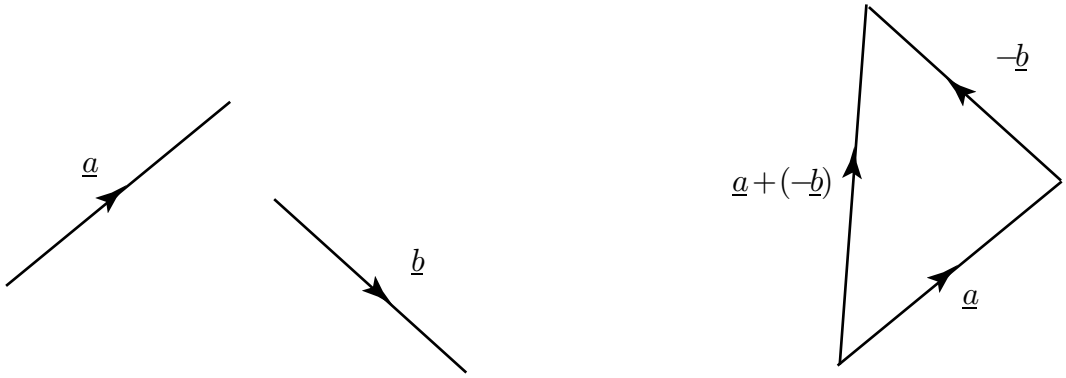


Figure 18: Subtraction of a vector is performed by adding a negative vector

Exercises

1. Vectors $\underline{p}$ and $\underline{q}$ represent two perpendicular sides of a square. Find vector expressions which represent the diagonals of the square.

2. In the rectangle $ABCD$, side AB is represented by the vector $\underline{p}$ and side BC is represented by the vector $\underline{q}$. State the physical significance of the vectors $\underline{p} - \underline{q}$ and $\underline{p} + \underline{q}$.

3. An object is positioned at the origin of a set of axes. Two forces act upon it. The first has magnitude 9 N and acts in the direction of the positive y axis. The second has magnitude 4 N and acts in the direction of the negative x axis. Calculate the magnitude and direction of the resultant force.

4. An object moves in the xy plane with a velocity of 15 m s^{-1} in a direction 48° above the positive x axis. Resolve this velocity into two components, one along the x axis and one along the y axis.

Answers

1. $\underline{p} + \underline{q}$, $\underline{q} - \underline{p}$. Acceptable answers are also $-(\underline{p} + \underline{q})$, $\underline{p} - \underline{q}$.

2. $\underline{p} + \underline{q}$ is the diagonal AC, $\underline{p} - \underline{q}$ is the diagonal DB.

3. Magnitude $\sqrt{97}$, at an angle 66° above the negative x axis.

4. 10.04 m s^{-1} along the x axis, and 11.15 m s^{-1} along the y axis.

5. Multiplying a vector by a scalar

If k is any positive scalar and $\underline{a}$ is a vector then $k\underline{a}$ is a vector in the same direction as $\underline{a}$ but k times as long. If k is negative, $k\underline{a}$ is a vector in the opposite direction to $\underline{a}$ and k times as long. See Figure 19. The vector $k\underline{a}$ is said to be a **scalar multiple** of $\underline{a}$.

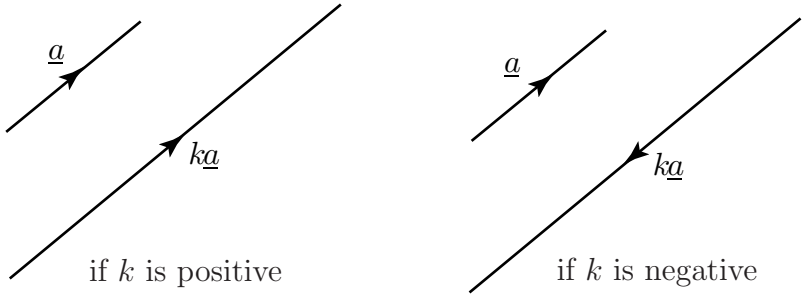


Figure 19: Multiplying a vector by a scalar

The vector $3\underline{a}$ is three times as long as $\underline{a}$ and has the same direction. The vector $\frac{1}{2}\underline{r}$ is in the same direction as $\underline{r}$ but is half as long. The vector $-4\underline{b}$ is in the opposite direction to $\underline{b}$ and four times as long.

For any scalars k and l, and any vectors $\underline{a}$ and $\underline{b}$, the following rules hold:

Key Point 2

$$k(\underline{a}+\underline{b}) = k\underline{a} + k\underline{b}$$

$$(k+l)\underline{a} = k\underline{a} + l\underline{a}$$

$$k(l)\underline{a} = (kl)\underline{a}$$

Using the rules given in Key Point 2, simplify the following:

(a) $3\underline{a} + 7\underline{a}$ (b) $2(7\underline{b})$ (c) $4\underline{q} + 4\underline{r}$

Your solution

Answer

(a) Using the second rule, $3\underline{a} + 7\underline{a}$ can be simplified to $(3+7)\underline{a} = 10\underline{a}$.

(b) Using the third rule $2(7\underline{b}) = (2 \times 7)\underline{b} = 14\underline{b}$.

(c) Using the first rule $4\underline{q} + 4\underline{r} = 4(\underline{q} + \underline{r})$.

Unit vectors

A vector which has a magnitude of 1 is called a **unit vector**. If $\underline{a}$ has magnitude 3, then a unit vector in the direction of $\underline{a}$ is $\frac{1}{3}\underline{a}$, as shown in Figure 20.

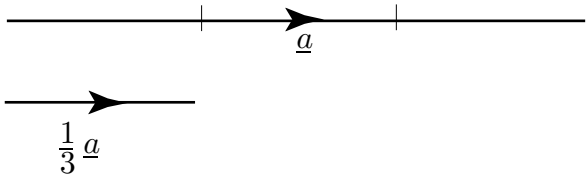


Figure 20: A unit vector has length one unit

A unit vector in the direction of a given vector is found by dividing the given vector by its magnitude: A unit vector in the direction of $\underline{a}$ is given the 'hat' symbol $\hat{\underline{a}}$.

Key Point 3

A unit vector can be found by dividing a vector by its modulus.

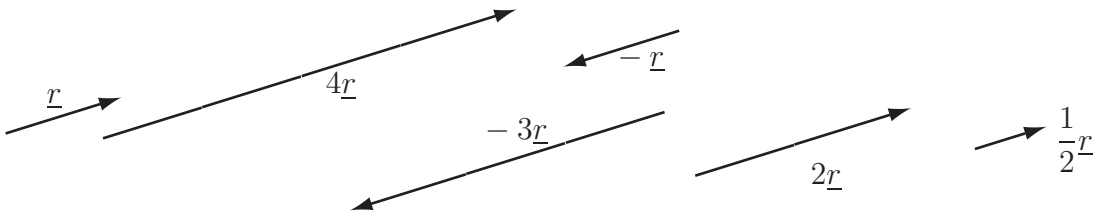
$$\hat{\underline{a}} = \frac{\underline{a}}{|\underline{a}|}$$

Exercises

1. Draw an arbitrary vector $\underline{r}$. On your diagram draw $2\underline{r}$, $4\underline{r}$, $-\underline{r}$, $-3\underline{r}$ and $\frac{1}{2}\underline{r}$.
2. In triangle OAB the point P divides AB internally in the ratio $m : n$. If $\overrightarrow{OA} = \underline{a}$ and $\overrightarrow{OB} = \underline{b}$ depict this on a diagram and then find an expression for $\overrightarrow{OP}$ in terms of $\underline{a}$ and $\underline{b}$.

Answers

1.



2. $\overrightarrow{OP} = \underline{a} + \dfrac{m}{m+n}(\underline{b} - \underline{a}) = \dfrac{n\underline{a} + m\underline{b}}{m+n}$.

Chapter Two: Vectors and Applications

This chapter contain 3 parts which is: vectors, vector components and the scalar product

Vectors

1.1 Vectors and Scalars

In everyday language speed and velocity mean the same thing, but in mathematics *speed* is a scalar and *velocity* is a vector.A *vector* is a quantity which is completely characterised by two things: its magnitude (or size) and its direction.

vectors

Example

- The velocity of a car heading North at 60 km/h.[1]
- A force exerted on an object.
- The magnetic field of the earth at a given place.

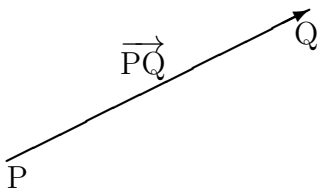
A *scalar* is a quantity which is completely characterised by its magnitude alone. Scalars are just numbers.

scalars

Example

- Length
- Area
- Volume
- Any number
- Mass
- Time
- Temperature

The simplest example of a vector is a *directed line segment* (or *arrow*). If P and Q are two points, then the directed line segment from P to Q is the straight line which begins at P and ends at Q. This vector is denoted by $\overrightarrow{PQ}$, and we say "the vector PQ".



[1]Velocity has both magnitude and direction, but speed has magnitude only. For example we would say the speed of the car is 60 km/h without referring to its direction.

The *magnitude or length* of $\overrightarrow{\text{PQ}}$ is the length of segment PQ, and is represented by the symbol $|\overrightarrow{\text{PQ}}|$. For example, if $\overrightarrow{\text{PQ}}$ is the velocity of a car, then $|\overrightarrow{\text{PQ}}|$ would be the speed of the car.

The *direction* of $\overrightarrow{\text{PQ}}$ is given by the arrowhead. A direction can be described in a number of ways.

• In navigation, direction is given by a compass bearing, e.g. NE, N45°E, N45° and 045° all refer to a bearing of 45° from north taken in a *clockwise* direction.

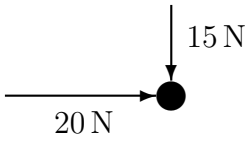
• In mathematics or engineering, direction is typically measured from a selected axis with angles taken in an *anticlockwise* direction.

All vectors can be represented by arrows or directed line segments.

vector diagram

Example

This *vector diagram* shows two forces acting on an object.
(Scale: 1cm = 10 Newtons)



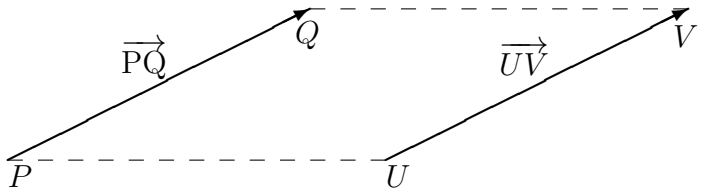
Definition 1.1.1

Two vectors are called **equal** *if they have the same length and direction.*

equal vectors

Example

The vectors $\overrightarrow{\text{PQ}}$ and $\overrightarrow{UV}$ below are equal as they have the same length and the same direction. It doesn't matter that the vectors are in different places, so long as their magnitudes and directions are the same. We write $\overrightarrow{\text{PQ}} = \overrightarrow{UV}$.

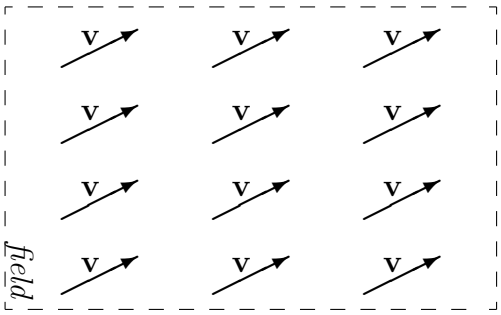


Vectors are usually represented in print by small boldface letters like **a**, **b**, **c**, ... , and in handwriting by $\underset{\sim}{a}$, $\underset{\sim}{b}$, $\underset{\sim}{c}$, ...

wind velocity

Example

Each of the vectors below represents the velocity **v** of the wind moving across a field. The wind has the same strength and direction everywhere on the field.

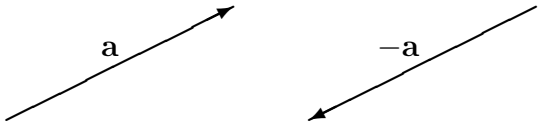


We think of negative numbers as being in the opposite direction to positive numbers on a numberline. The same is true for vectors.

Definition 1.1.2
The **negative** *of a vector* $\mathbf{a}$ *is a vector with the same length but the opposite direction. It is denoted by* $-\mathbf{a}$.

negative of a vector

Example

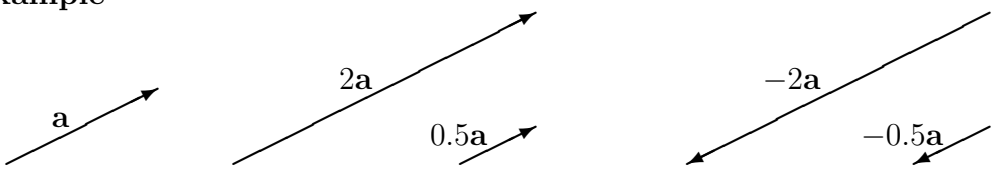


When a car doubles its velocity, then we think of it as travelling in the same direction but at twice the speed. The same is true when we double a vector. This is an example of *scalar multiplication*: multiplying a vector by the scalar 2.

Definition 1.1.3
If $\mathbf{a}$ *is a vector and* k *is a scalar (a number), then* $k\mathbf{a}$ *is the vector with*
- (if $k > 0$) *length* k *times the length of* $\mathbf{a}$ *and direction the same as* $\mathbf{a}$.
- (if $k < 0$) *length* $|k|$ *times the length of* $\mathbf{a}$ *and direction oppposite to* $\mathbf{a}$.[2]

scalar multiplication

Example



Exercise 1.1

1. After taking off, one plane flew East at 300 km/h and another flew 045° at 400 km/h.[3] Sketch these vectors using the scale 1cm = 100 km/h.

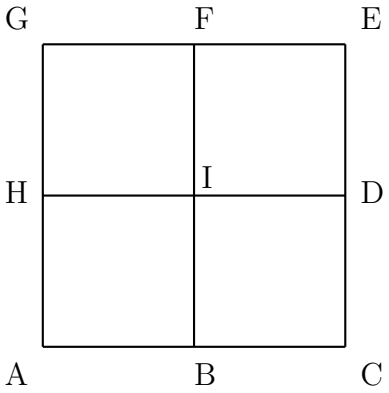
[2] The same symbol "| |" is used with vectors and scalars: $|k|$ is the absolute value of the number k,and $|\mathbf{a}|$ is the length of the vector $\mathbf{a}$.

[3] 45° from North in a clockwise direction.

2. Sketch the following vectors:
   (a) the velocity $\mathbf{v}$ of a car travelling south at 60 km/h.
   (b) $1.5\mathbf{v}$
   (c) $0.5\mathbf{v}$

3. A car was travelling with velocity $\mathbf{v}$, then 5 minutes later its velocity was $1.1\mathbf{v}$ and then 10 minutes later its velocity was $-\mathbf{v}$. What happened?

4. The figure below consists of 4 squares. $\overrightarrow{AB} = \mathbf{p}$ and $\overrightarrow{AH} = \mathbf{q}$.



   Which of the following is true?

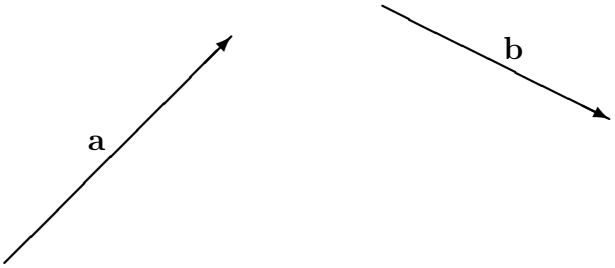
   (a) $\overrightarrow{FE} = \mathbf{p}$ (b) $|\mathbf{p}| = |\mathbf{q}|$ (c) $\overrightarrow{BC} = \mathbf{p}$ (d) $\overrightarrow{GF} = \mathbf{q}$
   (e) $\overrightarrow{GH} = \mathbf{q}$ (f) $\overrightarrow{GE} = \mathbf{2p}$ (g) $|\overrightarrow{HD}| = 2|\mathbf{q}|$ (h) $\overrightarrow{ED} = -\mathbf{q}$
   (i) $|\overrightarrow{FB}| = |\mathbf{2p}|$ (j) $\overrightarrow{GI} = \overrightarrow{FD}$ (k) $\overrightarrow{FG} = -\mathbf{p}$ (l) $\mathbf{p} = \mathbf{q}$

5. Is it true that $-(-\mathbf{a}) = \mathbf{a}$? Use a diagram to explain why.

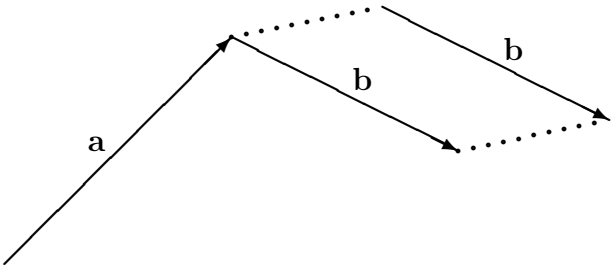
1.2 Adding and Subtracting Vectors

Although there are many different kinds of vectors, all behave in the same way. This is described in the definitions of vector addition and subtraction.

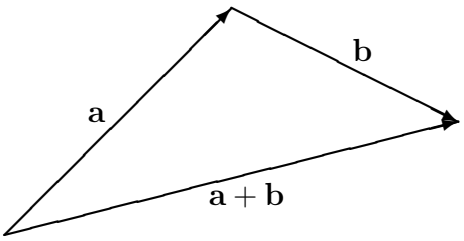
Definition 1.2.1
The **sum** $\mathbf{a} + \mathbf{b}$ *of vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *is the given by the triangle rule below.*



Step 1: Shift $\mathbf{b}$ *without changing its length or direction, until its tail just touchs the head of* $\mathbf{a}$.



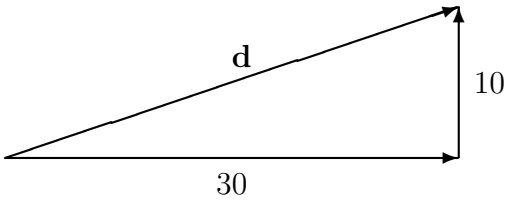
Step 2: The vector $\mathbf{a} + \mathbf{b}$ *is equal to the vector that begins at the tail of* $\mathbf{a}$ *and ends at the head of* $\mathbf{b}$.



This method is called the *triangle method* or the *head-to-tail method* of addition.

resultant displacement

Example
We can describe how an object is *displaced* by giving the distance and direction of the displacement. In the diagram below an object is first displaced towards the east by 30 km, and then north by 10 km.

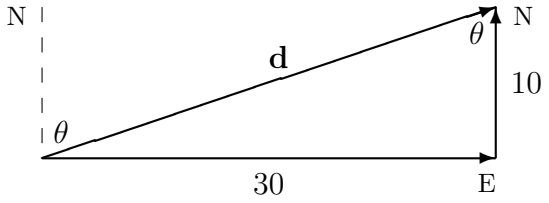


The vector **d** shows the object's *resultant* displacement. This means the displacement resulting from combining the actions of the vectors.

The length or magnitude $|\mathbf{d}|$ of the vector **d** can be estimated from accurately scaled vector diagrams or by using trigonometry.

Example (continued)

finding length direction



The length of **d** is

$$|\mathbf{d}| = \sqrt{30^2 + 10^2} = \sqrt{1000} = 31.6\,\text{km}.$$

The direction of **d** is Nθ°, where

$$\tan\theta = \frac{30}{10} \Longrightarrow \theta = 71.6^\circ.$$

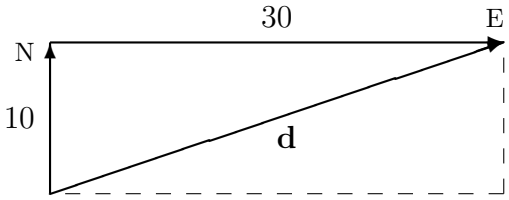
That is, the direction is the compass bearing of N71.6° (which is measured clockwise from North).

The resultant vector **d** *does not depend on the order in which vectors are added in the head-to-tail method.*

Example

order of addition

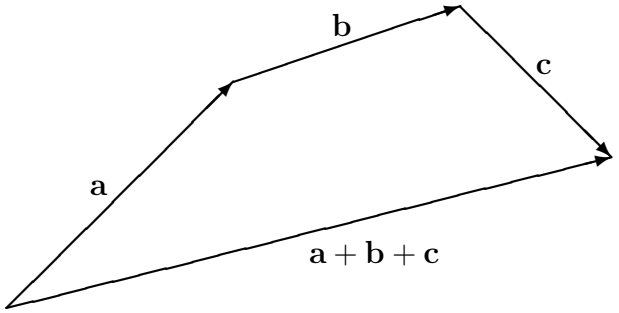
If the object is first displaced towards the north by 10 km, then east by 30 km, the resultant displacement **d** is still the same as in the example above.



The head-to-tail method can be used for adding three or more vectors. It does not matter in which order these are combined.

adding three vectors

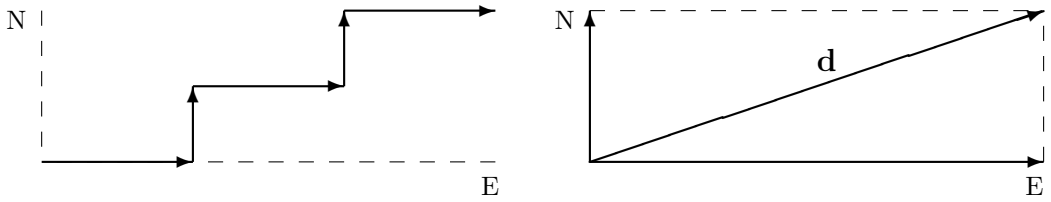
Example



rearranging vectors

Example

A plane flew 30 km east and 10 km north in a zig-zag fashion. The diagrams below show that it doesn't matter how we rearrange and combine the zigs and zags, the total displacement to the east is 30 km, and the total displacement to the west is 10 km. The resultant displacement is always the same.

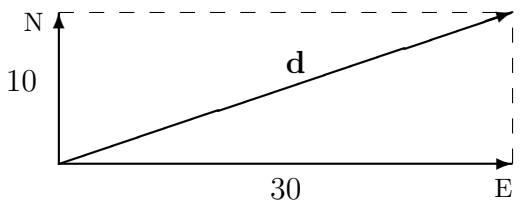


In the physical world many vectors interact simultaneously. For example, a plane may be displaced both towards the east and towards the north at the same time. In this case the resultant vector can still be found by adding the the individual vectors that produce the action. The example above suggests why if we imagine that the simultaneous displacements correspond to a large number of zig-zags.

simultaneous displacements

Example (continued)

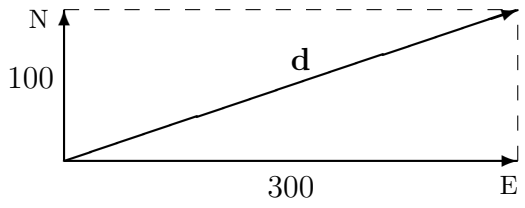
If a plane travels 30 km to the east, and is *simultaneously* blown off course by 10 km to the north, then we can think of the resultant displacement **d** as consisting of two parts, (a) the displacement 30 km east and (b) the displacement 20 km north.



Example

resultant velocity

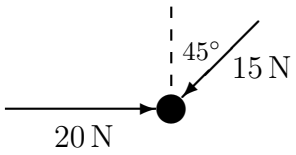
If a plane travels at 300 km/h towards the east, and it is *simultaneously* blown off course by a 100 km/h wind from the south, then we can think of its resultant velocity $\mathbf{v}$ as consisting of two parts, the first a velocity of 300 km/h to the east and the second a 100 km/h to the north. The resultant velocity - both its length and direction - can be calculated from knowing these two velocities.



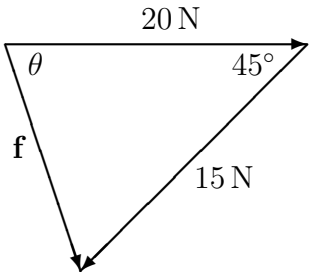
Example

resultant force

The vector diagram shows two forces acting on a ball bearing. What is the resultant force?[4]



Answer



The length of $\mathbf{f}$ can be found using the cosine rule.

$$\begin{aligned} |\mathbf{f}|^2 &= 20^2 + 15^2 - 2 \times 20 \times 15 \times \cos 45^\circ = 200.74 \\ |\mathbf{f}| &= 14.17 \end{aligned}$$

The angle θ can be found from the sine rule.

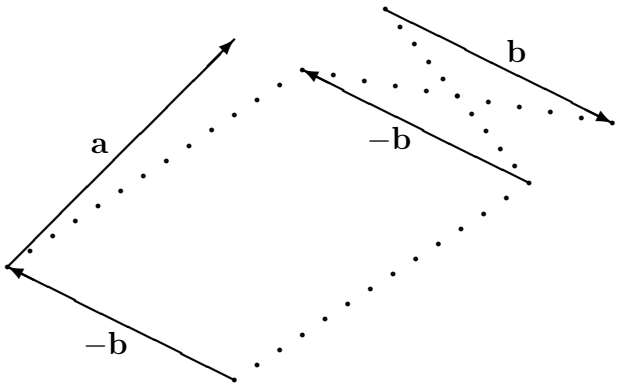
$$\begin{aligned} \frac{\sin\theta}{15} &= \frac{\sin 45}{14.17} \\ \sin\theta &= 0.7485 \\ \theta &= 48.46 \end{aligned}$$

The resultant force $\mathbf{f}$ has length 14.17 N and direction -48.46° relative to the horizontal.

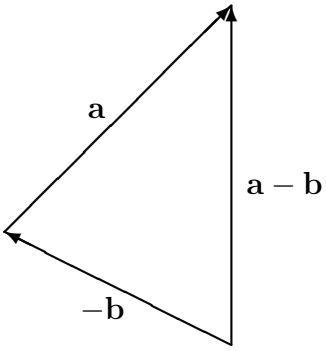
[4]The resultant force can be thought of as a single force which is equivalent to the two forces shown in the diagram

Definition 1.2.2
The difference $\mathbf{a} - \mathbf{b}$ *of two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *is the sum of* $\mathbf{a}$ *and* $-\mathbf{b}$.

Step 1: Construct $-\mathbf{b}$*, then shift it until its head just touches the tail of* $\mathbf{a}$.



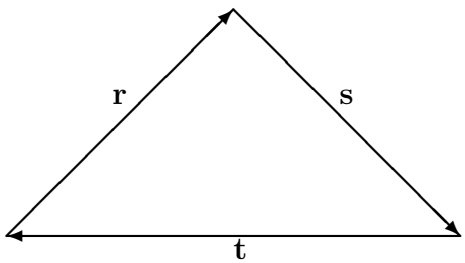
Step 2: The vector $\mathbf{a} - \mathbf{b}$ *is equal to the vector that begins at the tail of* $-\mathbf{b}$ *and ends at the head of* $\mathbf{a}$.



Definition 1.2.3
The zero vector $\mathbf{0}$ *has length 0. Its direction is not defined.*

a vector equation

Example
The vectors $\mathbf{r}$, $\mathbf{s}$ and $\mathbf{t}$ below satisfy the vector equation $\mathbf{r} + \mathbf{s} + \mathbf{t} = \mathbf{0}$.



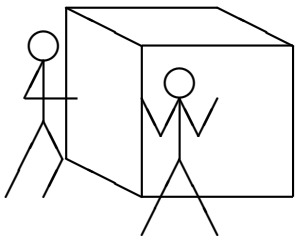
Exercise 1.2

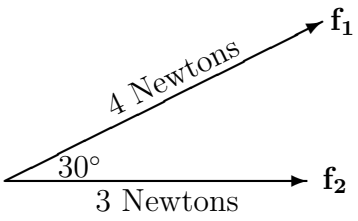
NOTE: These exercises use the rules of trigonometry, which are covered in MathsStart Topics 5 and 6.

1. A plane is flying NE at its maximum velocity of 500 km/h according to its instruments. However there is a northerly[5] wind of 100 km.
   (a) Sketch this situation on a vector diagram.
   (b) Calculate the resultant velocity of the plane.

2. A yacht is cruising at 15 knots on a compass bearing of N124°. Find the actual speed and direction of the yacht if there is also an ocean current of 5 knots in the direction N58°.

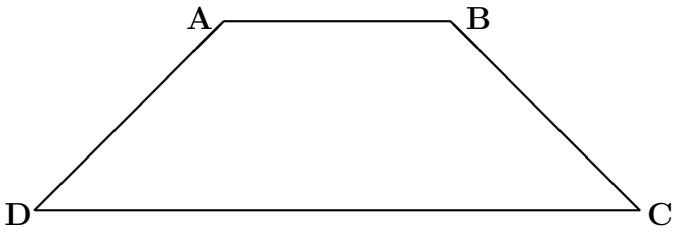
3. Two students push a heavy object. The vector diagram below shows the forces on that object.





   Find the magnitude and direction of the resultant force.

4. The four points A, B, C, D are vertices of the trapezium.

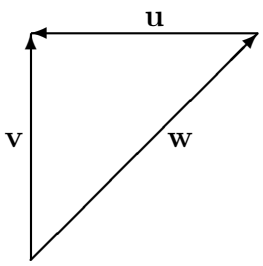


   Simplify the following vector expressions as much as possible:
   (a) $\overrightarrow{DA} + \overrightarrow{AB} + \overrightarrow{BC}$
   (b) $\overrightarrow{DB} - \overrightarrow{AB}$
   (c) $\overrightarrow{CA} - \overrightarrow{DA} + \overrightarrow{DB}$
   (d) $\overrightarrow{DA} + \overrightarrow{AD}$

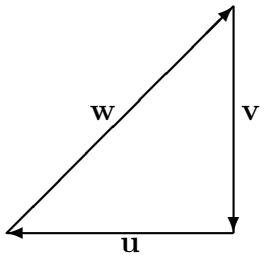
[5]Note that a northerly wind blows *from* the North!

5. Construct vector equations for $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ if:

(i)



(ii)



6. A pilot wants to fly due east at a speed of 400 km/h, however there is a 50km/h wind towards the SE. What direction must the pilot fly in, and at what speed?

7. Use a diagram to show that $\mathbf{a} - \mathbf{a} = \mathbf{0}$ for any vector $\mathbf{a}$.

1.3 Vector Algebra

We can perform a limited amount of algebra on vectors: adding, subtracting, scalar multiplication, simplfying by collecting and combining like terms, multiplying out brackets, and solving linear equations. These techniques contribute to the usefulness of vectors in applications.

collecting & combining like terms

Example
In a vector expression like $\mathbf{2a} + \mathbf{b} + \mathbf{a} + \mathbf{3b} - \mathbf{4a}$, the terms $\mathbf{2a}$, $\mathbf{a}$ and $-\mathbf{4a}$ are called *like terms* because each is a scalar multiple of the vector $\mathbf{a}$. We can combine like terms by adding their coefficients.

$$\mathbf{2a} + \mathbf{a} - \mathbf{4a} = (\mathbf{2} + \mathbf{1} - \mathbf{4})\mathbf{a} = -\mathbf{a}$$

In a similar way we can combine the multiples of $\mathbf{b}$.

We simplify $\mathbf{2a} + \mathbf{b} + \mathbf{a} + \mathbf{3b} - \mathbf{4a}$ by collecting and combining like terms, in exactly the same way as we do in ordinary algebra:

$$\mathbf{2a} + \mathbf{b} + \mathbf{a} + \mathbf{3b} - \mathbf{4a} = -\mathbf{a} + \mathbf{3b}.$$

expanding brackets

Example
Vector expressions may also contain brackets such as $\mathbf{2}(\mathbf{a} + \mathbf{3b})$. These can be multiplied out as in ordinary algebra.

$$\mathbf{2a} + \mathbf{b} - \mathbf{2}(\mathbf{a} + \mathbf{3b}) = \mathbf{2a} + \mathbf{b} - \mathbf{2a} - \mathbf{6b} = -\mathbf{5b}$$

solving vector equations

Example
We can solve vector equations in *almost* the same way that we solve ordinary linear equations[6], the main difference being that it is customary to *multiply but not divide* vectors by scalars.

Solve the equation $\mathbf{2x} + \mathbf{a} - \mathbf{b} + \mathbf{c} = \mathbf{2a} - \mathbf{3b} + \mathbf{c}$ for the unknown vector $\mathbf{x}$.

Answer

$$\begin{aligned} \mathbf{2x} + \mathbf{a} - \mathbf{b} + \mathbf{c} &= \mathbf{2a} - \mathbf{3b} + \mathbf{c} \\ \mathbf{2x} &= \mathbf{a} - \mathbf{2b} \qquad \text{... now } \textit{multiply} \text{ both sides by } \tfrac{1}{2}. \\ \mathbf{x} &= \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{a} - \mathbf{2b}) \text{ or } \frac{\mathbf{1}}{\mathbf{2}}\mathbf{a} - \mathbf{b} \end{aligned}$$

As you can see, the methods we use for simplifying linear expressions in ordinary algebra carry intuitively over to vectors. *However*, vectors are different from numbers, and we should confirm that what we are doing is valid.

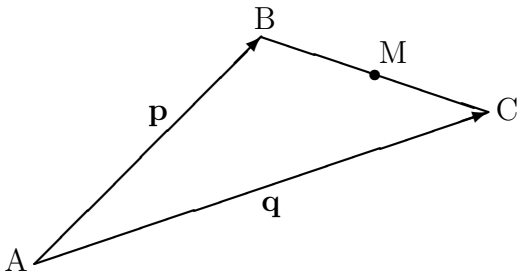
[6] A *linear* equation is an equation in which the unknown only occurs to the first power.

1.4 Geometric Problems

Vector algebra can be used to solve geometric problems.

midpoint formula

Example
In the vector diagram M is the midpoint of BC. Vector algebra can be used to find an expression for $\overrightarrow{AM}$ in terms of $\mathbf{p}$ and $\mathbf{q}$.



The diagram shows that

$$\overrightarrow{AM} = \overrightarrow{AB} + \overrightarrow{BM} = \mathbf{p} + \overrightarrow{BM}$$

As M is the midpoint of BC, you can see that

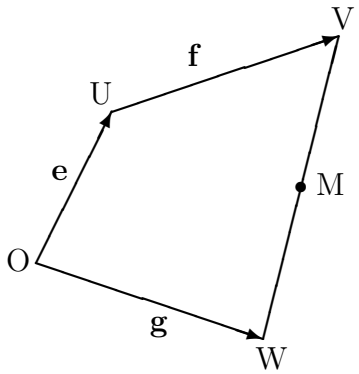
$$\overrightarrow{BM} = \frac{1}{2}\overrightarrow{BC} = \frac{1}{2}(\mathbf{q} - \mathbf{p})$$

Substituting this into the previous equation:

$$\begin{aligned} \overrightarrow{AM} &= \mathbf{p} + \frac{1}{2}(\mathbf{q} - \mathbf{p}) \\ &= \frac{1}{2}(\mathbf{p} + \mathbf{q}) \end{aligned}$$

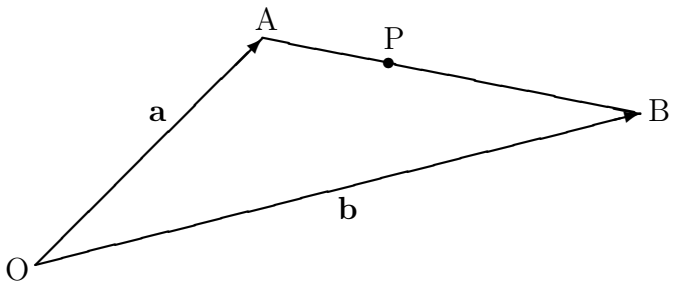
Exercise 1.4

1. In the diagram M is the midpoint of VW.



Find expressions for $\overrightarrow{OV}$, $\overrightarrow{VW}$, $\overrightarrow{UM}$ and $\overrightarrow{OM}$ in terms of $\mathbf{e}$, $\mathbf{f}$, $\mathbf{g}$.

2. In the diagram P is divides AB in the ratio 1:2. Find an expression for $\overrightarrow{OP}$ in terms of **a** and **b**.

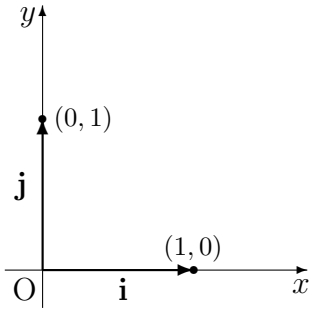


Vector Components

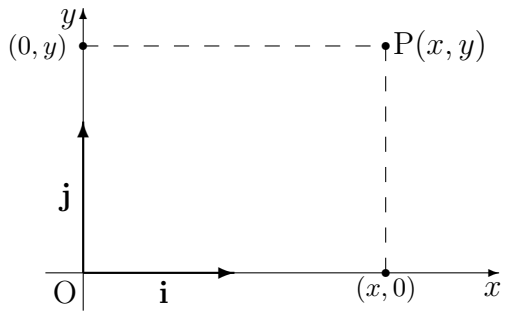
2.1 Vectors and Co-ordinates

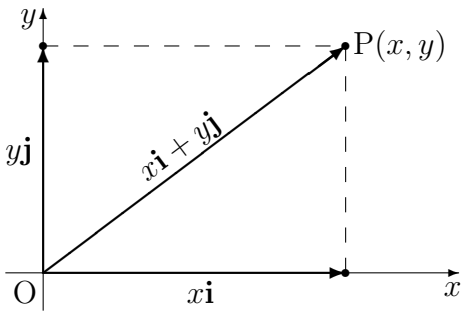
If a Cartesian co-ordinate system was added to a vector diagram, then co-ordinates could be assigned to the top and tail of each vector.We could then describe vectors in terms of these co-ordinates, and use this information in calculations.

It is convenient to introduce vectors **i** and **j** on the x- and y-axes, each with tail at the origin and pointing in the positive x- and y-directions, and each with length 1. These are called *unit vectors* as they have length 1. When talking about them, we just call them "i" and "j".



We can use co-ordinates to describe vectors ...and we can use vectors to locate points on the Cartesian plane.





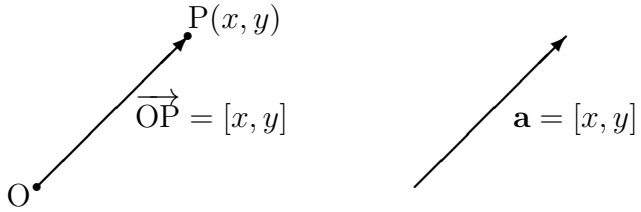
Definition 2.1.1
If $P(x, y)$ is a point on the Cartesian plane, and if O is the origin then $\overrightarrow{OP} = x\mathbf{i} + y\mathbf{j}$ is called the **position vector** *of the point P.*

The notation $[x, y]$ is commonly used as an abbreviation for the vector $x\mathbf{i} + y\mathbf{j}$. The numbers x and y are called the *vector components* of $x\mathbf{i}+y\mathbf{j}$, x being the x-component and y being the y-component.

vector components

Example
(a) $\mathbf{i} = [1, 0]$ (b) $\mathbf{j} = [0, 1]$ (c) $\mathbf{2}i - 3j = [2, -3]$ (d) $\mathbf{0} = [0, 0]$

If $\mathbf{a}$ is a vector which is equal to position vector $\overrightarrow{\mathrm{OP}} = [x, y]$, then we can also write $\mathbf{a} = [x, y]$, and also say that the numbers x and y are the vector components of $\mathbf{a}$.

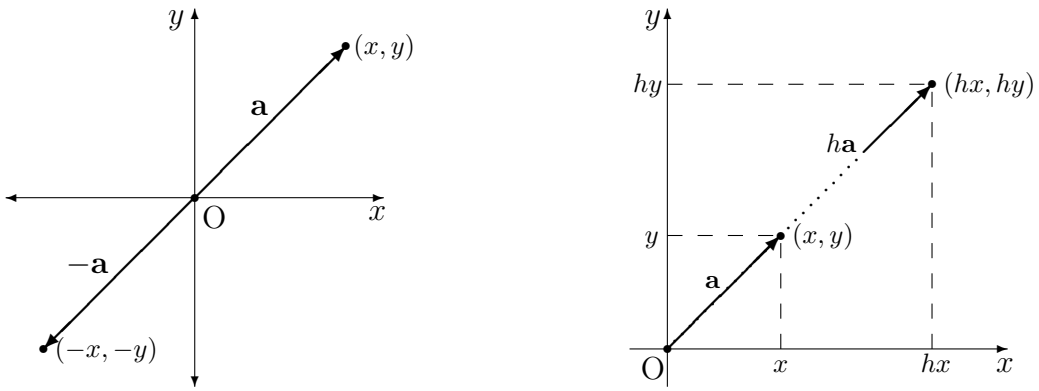


Working with components is very intuitive.

Properties of Components

- If $\mathbf{a} = [x, y]$, then $-\mathbf{a} = [-x, -y]$
- If $\mathbf{a} = [x, y]$ and h is a scalar, then $h\mathbf{a} = [hx, hy]$
- If $\mathbf{a} = [x_1, y_1]$ and $\mathbf{b} = [x_2, y_2]$, then $\mathbf{a} \pm \mathbf{b} = [x_1 \pm x_2, y_1 \pm y_2]$

The first two properties are illustrated below for position vectors, but are true for *all* vectors.



The three properties can be *proved* using vector algebra.

Example

adding vector components

Show that if $\mathbf{a} = [x_1, y_1]$ and $\mathbf{b} = [x_2, y_2]$, prove that $\mathbf{a} + \mathbf{b} = [x_1 + x_2, y_1 + y_2]$.

Answer

$$\begin{aligned} \mathbf{a} + \mathbf{b} &= [x_1, y_1] + [x_2, y_2] \\ &= (x_1\mathbf{i} + y_1\mathbf{j}) + (x_2\mathbf{i} + y_2\mathbf{j}) \\ &= (x_1 + x_2)\mathbf{i} + (y_1 + y_2)\mathbf{j}) \\ &= [x_1 + x_2, y_1 + y_2] \end{aligned}$$

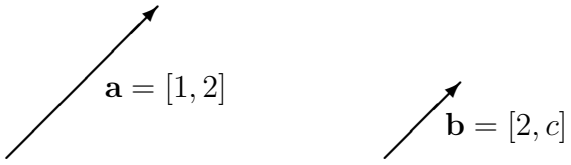
Example

parallel vectors

Find c if $\mathbf{a} = [1, 2]$ and $\mathbf{b} = [2, c]$ are parallel.

Answer

Two vectors can only be parallel if one is a scalar multiple of the other.



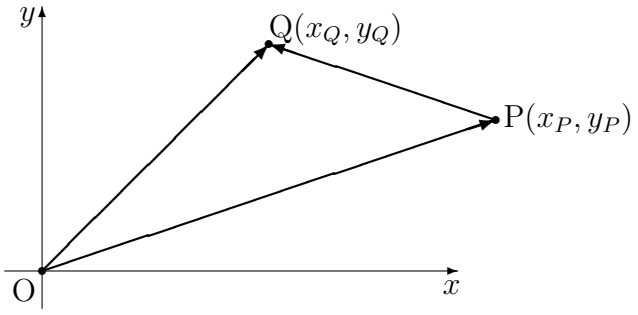
If $\mathbf{a}$ and $\mathbf{b}$ are parallel, then $\mathbf{a} = h\mathbf{b}$ for some scalar h.

Put	$[1, 2] = h[2, c]$,
then	$2h = 1$ and $ch = 2$,
so	$h = \frac{1}{2}$ and $c = 4$.

Theorem[1] If $\mathrm{P}(x_P, y_P)$ and $\mathrm{Q}(x_Q, y_Q)$ are any two points in a plane, then

$$\overrightarrow{\mathrm{PQ}} = [x_Q - x_P, y_Q - y_P].$$

Proof.



[1] A *theorem* is an important or useful result.

It can be seen from the diagram that the position vectors $\overrightarrow{\text{OP}}$ and $\overrightarrow{\text{OQ}}$ are given by

$$\begin{aligned} \overrightarrow{\text{OP}} &= [x_P, y_P] = [x_P\mathbf{i} + y_P\mathbf{j} \\ \text{and } \overrightarrow{\text{OQ}} &= [x_Q, y_Q] = [x_Q\mathbf{i} + y_Q\mathbf{j}]. \end{aligned}$$

$$\begin{aligned} \text{So } \overrightarrow{\text{PQ}} &= \overrightarrow{\text{OQ}} - \overrightarrow{\text{OP}} \\ &= x_Q\mathbf{i} + y_Q\mathbf{j} - ([x_P\mathbf{i} + y_P\mathbf{j}]) \\ &= (x_Q - x_P)\mathbf{i} + (y_Q - y_P)\mathbf{j} \\ &= [x_Q - x_P, y_Q - y_P]. \end{aligned}$$

In words: the components of a directed line segment are the co-ordinates of the top minus the co-ordinates of the tail.

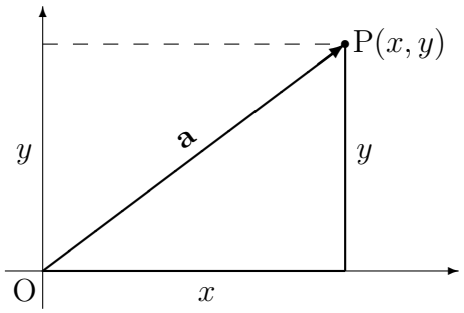
Exercise 2.1

1. If $\mathbf{a} = [2, -3]$ and $\mathbf{b} = [3, 4]$, find the vector components of
   (i) $3\mathbf{a} + 4\mathbf{b}$
   (ii) $2\mathbf{b} - \mathbf{a}$
2. Find c if $\mathbf{a} = [5, -2]$ and $\mathbf{b} = [10, c]$ are parallel.
3. If $\mathbf{a} = [2, -1]$, $\mathbf{b} = [3, 4]$ and $\mathbf{c} = [p, 15]$, find p so that $\mathbf{c}$ is parallel to $\mathbf{b} - \mathbf{a}$.
4. If P$(1, -2)$, Q$(-3, 1)$, R$(2, 4)$ and S$(-1, 6)$ are points the Cartesian plane, find the components of the vector $\overrightarrow{\text{PQ}} + 2\overrightarrow{\text{RS}}$.

2.2 Formula for Length

The magnitude or length of any vector can be found from its components using Pythagoras' theorem.

The diagram below shows that the length of $\mathbf{a} = [x, y]$ is $|\mathbf{a}| = \sqrt{x^2 + y^2}$.



Theorem If $\mathbf{a}$ is a vector and h is a scalar, then $|h\mathbf{a}| = |h||\mathbf{a}|$.

Proof. Let $\mathbf{a} = [x, y]$, then $h\mathbf{a} = [hx, hy]$ and

$$|h\mathbf{a}| = \sqrt{(hx)^2 + (hy)^2} = h\sqrt{(x^2 + y^2} = |h||\mathbf{a}|.$$

Finding a unit vector

Corollary If $\mathbf{a}$ is a vector, then $\frac{1}{|\mathbf{a}|}\mathbf{a}$ has length 1 (ie. it is a unit vector having the same direction as $\mathbf{a}$).[2,3]

Proof. The theorem shows that the length of $\frac{1}{|\mathbf{a}|}\mathbf{a}$ is $\left|\frac{1}{|\mathbf{a}|}\mathbf{a}\right| = \frac{1}{|\mathbf{a}|}|\mathbf{a}| = 1$.

Exercise 2.2

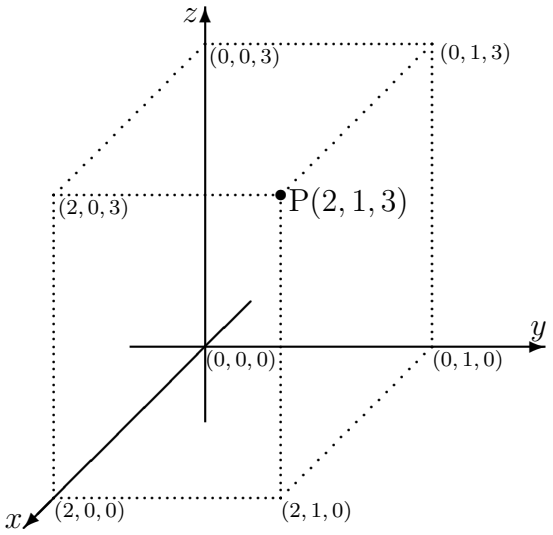
1. If $\mathbf{a} = [1, 3]$ and $\mathbf{b} = [2, -1]$, find $|\mathbf{a} - 2\mathbf{b}|$.
2. If $\mathbf{a} = [1, 4]$, $\mathbf{c} = [2, 1]$ and $\mathbf{a} + \mathbf{b} = \mathbf{c}$, find $|\mathbf{b}|$.
3. Find unit vectors having the (i) same direction and (ii) opposie direction to (i) $\mathbf{a} = [3, -4]$ (ii) $\mathbf{b} = [-5, 12]$ (iii) $\mathbf{c} = [-6, 8]$ (iv) $\mathbf{O} = [0, 0]$.
4. If $\mathrm{P}(1, -1)$, $\mathrm{Q}(-2, 1)$, $\mathrm{R}(2, 4)$ and $\mathrm{S}(-1, 5)$ are points, find the unit vector parallel to $\overrightarrow{\mathrm{PQ}} + 2\overrightarrow{\mathrm{RS}}$.

[2] A *corollary* is a deduction from a theorem.

[3] Unit vectors like this are used in astronomy, where the direction of a star is may be known but its distance from earth is not.

2.3 Vectors in Three Dimensions

All of the previous ideas can be extended to three dimensions. The diagram below shows a three dimensional co-ordinate system.



If you arrange your thumb and the first two fingers on your right hand in the same directions as the three axes in the diagram, with the thumb pointing in the z-direction, then first finger point will point in the x-direction and the second finger will point in the y-direction. When a screw is screwed in the z-direction, your right hand will rotate (anti-clockwise) from the x-axis towards the y-axis.

The point P has co-ordinates $(2, 1, 3)$. It can be reached by moving 2 units along the x-axis from the origin $(0, 0, 0)$, then moving 1 unit parallel to the y-axis, then moving 3 units parallel to the z-axis.

The *projection* of point $P(2, 1, 3)$ onto the (x, y)-plane is $(2, 1, 0)$, the projection onto the (y, z)-plane is $(0, 1, 3)$, and the projection onto the (z, x)-plane is $(2, 0, 3)$.

Example
In three dimensions, **i**, **j** and **k** represent the unit vectors which have tails at the origin and which point along the x-, y- and z-axes in a positive direction.

The position vector of the point P$(2, 1, 3)$ above is $\overrightarrow{\text{PQ}} = 2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$.

$\overrightarrow{\text{PQ}}$ has components $[2, 1, 3]$ and length $|\overrightarrow{\text{PQ}}| = \sqrt{2^2 + 1^2 + 3^2} = \sqrt{14}$.

In general, the length of the three dimensional vector $[x, y, z]$ is $\sqrt{x^2 + y^2 + z^2}$.

Exercise 2.3

1. Draw a three dimensional co-ordinate system and show the points:
   (i) $(3, 0, 0)$ (ii) $(3, 2, 0)$ (iii) $(3, 2, 1)$ (iv) $(-1, -1, -1)$
2. How far is $P(3, 4, 5)$ from the origin?
3. If $\mathbf{a} = [1, 3, 2]$ and $\mathbf{b} = [2, -1, 4]$, find $|\mathbf{a} - 2\mathbf{b}|$.
4. If $\mathbf{a} = [1, 0, 4]$, $\mathbf{c} = [2, 1, 0]$ and $\mathbf{a} + 2\mathbf{b} = \mathbf{c}$, find $|\mathbf{b}|$.
5. Find a unit vector having the same direction as $\mathbf{a} = [3, -4, 5]$
6. A rectangular prism (box) has sides 2m, 3m and 4m long. What is the maximum distance between 2 corners?

The Scalar Product

3.1 Scalar Products and Angles

Scalar products enable the angle between two vectors to be calculated.

Definition 3.1.1

1. *If* $\mathbf{a} = [a_1, a_2]$ *and* $\mathbf{b} = [b_1, b_2]$, *then the* scalar product *of* $\mathbf{a}$ *and* $\mathbf{b}$ *is*

$$\mathbf{a} \bullet \mathbf{b} = a_1 b_1 + a_2 b_2.$$

2. *If* $\mathbf{a} = [a_1, a_2, a_3]$ *and* $\mathbf{b} = [b_1, b_2, b_3]$, *then the* scalar product *of* $\mathbf{a}$ *and* $\mathbf{b}$ *is*

$$\mathbf{a} \bullet \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

Scalar products are also called *dot products*.

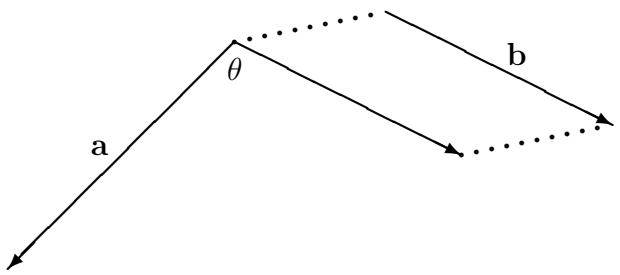
scalar or dot products

Example

(a) If $\mathbf{a} = [1, 4]$ and $\mathbf{b} = [2, -3]$, then $\mathbf{a} \bullet \mathbf{b} = 1 \times 2 + 4 \times -3 = -10$.

(b) If $\mathbf{u} = [1, 1, 2]$ and $\mathbf{v} = [1, 1, -1]$, then $\mathbf{u} \bullet \mathbf{v} = 1 \times 1 + 1 \times 1 + 2 \times -1 = 0$.

The angle between two position vectors is the *smallest* angle they make at the origin. If the vectors are not position vectors, then we can think of the angle between them as being the angle between their equivalent position vectors. Alternatively, it can be thought of as the smallest angle between them if they are shifted so that their tails touch.



The scalar product is important because of the following theorem.

Theorem If θ is the angle between two vectors $\mathbf{a}$ and $\mathbf{b}$, then

$$\mathbf{a} \bullet \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta.$$

finding an angle

Example

Find the angle between the vectors $\mathbf{a} = [3, 2]$ and $\mathbf{b} = [2, 1]$

Answer

The dot product is $\mathbf{a} \bullet \mathbf{b} = 3 \times 2 + 2 \times 1 = 8$.

The lengths are $|\mathbf{a}| = \sqrt{3^2 + 2^2} = \sqrt{13}$ and $|\mathbf{b}| = \sqrt{2^2 + 1^2} = \sqrt{5}$.

The angle is given by

$$\cos\theta = \frac{\mathbf{a} \bullet \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \frac{8}{\sqrt{13}\sqrt{5}} = \frac{8}{\sqrt{65}}.$$

So $\theta = 7.12^o$

perpendicular vectors

Example

Show that the position vectors [1, 1, 2] and [1, 1, -1] are perpendicular.

Answer

The dot product is $\mathbf{a} \bullet \mathbf{b} = 1 \times 1 + 1 \times 1 + 2 \times -1 = 0$.

This shows that $\cos\theta = 0$, so the vectors are perpendicular.

Exercise 3.1

1. Find the angle between $[1, 2]$ and $[3, -3]$.
2. Find the angle that points $P(1, 1)$ and $Q(1, 2)$ subtend at the origin.
3. Find the angle at which the line between $P(1, 1)$ and $Q(2, 3)$ intersects the x-axis. (Hint: Use the unit vector $\mathbf{i}$ in your calculations)
4. A cube has sides 1 m long. If three of the sides are taken to be the x-, y- and z- axes, find the angle between z-axis and the diagonal running between the origin and the furthest edge.
5. A rectangular box has dimensions 1m, 2m and km. What should the value of k be if the angle between longest diagonal and the longest edge of the box needs to be 20^o?

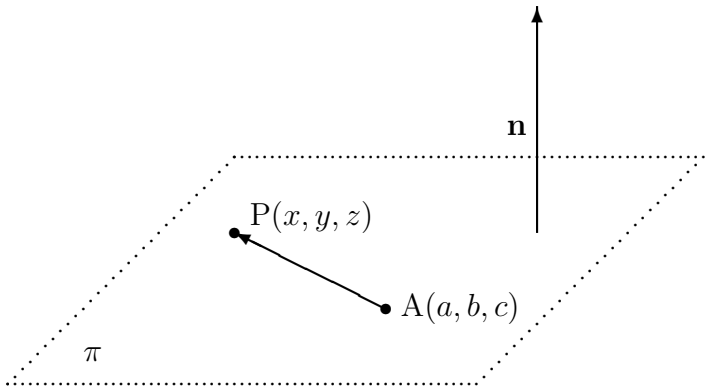
3.2 Planes in 3-dimensions

A vector is said to be normal to a plane if it is perpendicular to every vector that lies in the plane.

The equation of a plane can be found from:

- a point $\mathrm{A}(a, b, c)$ on the plane.
- a normal vector $\mathbf{n} = [l, m, n]$ to the plane

In the diagram below, (x, y, z) is a general point on the plane π.



As $AP = [x-a, y-b, z-c]$ and $\mathbf{n} = [l, m, n]$ are perpendicular, their scalar product is zero and

$$l(x-a) + m(y-b) + n(z-c) = 0$$

The equation of π is $l(x-a) + m(y-b) + n(z-c) = 0$.

equation of a plane

Example
Find the equation of the plane with normal vector $[1, 0, -2]$ that passes through the point $(1, 1, 3)$, and find one point on the plane.

Answer

i. The equation is $1(x-1) + 0(y-1) + -2(z-3) = 0$, that is $x - 2z + 5 = 0$
ii. Let $z = 0$, then $x = -5$. the equation imposes no restrictions on the value of y, so take $y = 1$. One point is $(-5, 1, 0)$.

normal to a plane

Example
Find a normal vector to the plane $x + 2y - z = 4$

Answer

A normal vector is [1, 2, -1].

Exercise 3.2

1. Find the equation of a plane with normal vector $[2, 3, -1]$ that passes through $(1, 1, 1)$.
2. $A(1, 2, -1)$ and $B(2, 1, -1)$ are two points. Find the equation of the plane which passes through A and is normal to AB.
3. Find a unit vector normal to the plane x + y + z = 3.
4. Find the equation of the plane perpendicular to the z-axis and passing through the point $(1, -2, 5)$

Appendix A

Answers

Exercise 1.1

1. check with tutor
2. check with tutor
3. After 5 minutes, it increased its velocity by 10 percent, then 10 minutes later it turned in the opposite direction and moved at the same velocity it started with.
4. (d), (e) and (l) are false
5. check with tutor

Exercise 1.2

1. (a) check with tutor
   (b) Magnitude $= \sqrt{500^2 + 100^2 - 2 \times 500 \times 100 \times \cos 45} = 435.1$,
   Direction $= 45° + \theta$ from North, where

$$\frac{\sin\theta}{100} = \frac{\sin 45}{435.1} \Rightarrow \theta = 9.353^o$$

   That is, the direction is a compass bearing of N54.353°.

2. (a) Magnitude $= \sqrt{5^2 + 15^2 - 2 \times 5 \times 15 \times \cos 114} = 17.64$
   (b) Direction = N109.0°
3. (a) Magnitude $= \sqrt{3^2 + 4^2 - 2 \times 3 \times 4 \times \cos 150} = 6.766$
   (b) Direction = 17.19° above horizontal.
4. (a) $\overrightarrow{DC}$ (b) $\overrightarrow{DA}$ (c) $\overrightarrow{CB}$ (d) $\mathbf{0}$
5. (a) $\mathbf{u} - \mathbf{v} + \mathbf{w} = \mathbf{0}$ (b) $\mathbf{u} + \mathbf{w} + \mathbf{v} = \mathbf{0}$

6. (a) Magnitude = 436.9 km/h
   (b) Direction = 4.6° S of E (or a compass bearing of N94.6°).

7. check with tutor

Exercise 1.4

1. (a) $\overrightarrow{\text{OV}} = \mathbf{e} + \mathbf{f}$
   (b) $\overrightarrow{\text{VW}} = -\mathbf{f} - \mathbf{e} + \mathbf{g}$
   (c) $\overrightarrow{\text{UM}} = \mathbf{f} + \frac{1}{2}(-\mathbf{f} - \mathbf{e} + \mathbf{g}) = \frac{1}{2}\mathbf{f} - \frac{1}{2}\mathbf{e} + \frac{1}{2}\mathbf{g}$
   (d) $\overrightarrow{\text{OM}} = \mathbf{e} + \mathbf{f} + \frac{1}{2}(-\mathbf{f} - \mathbf{e} + \mathbf{g}) = \frac{1}{2}\mathbf{e} + \frac{1}{2}\mathbf{f} + \frac{1}{2}\mathbf{g}$

2. $\frac{2}{3}\mathbf{a} + \frac{1}{3}\mathbf{b}$

Exercise 2.1

1. (i) $[18, 7]$ (ii) $[4, 11]$

2. $c = -4$

3. $p = 3$

4. $[-10, 7]$

Exercise 2.2

1. $\sqrt{34}$

2. $\sqrt{10}$

3. (i) $\pm\frac{1}{5}[3, -4]$
   (ii) $\pm\frac{1}{13}[5, -12]$
   (iii) $\pm\frac{1}{5}[-3, 4]$
   (iv) the direction is not defined.

4. $\frac{1}{\sqrt{97}}[-9, 4]$

Exercise 2.3

1. check with tutor

2. $\sqrt{50}$

3. $\sqrt{70}$

4. $\sqrt{\frac{9}{2}}$

5. $\frac{1}{\sqrt{50}}[3, -4, 5]$

6. $\sqrt{29}$ m

Exercise 3.1

1. 108.4°
2. 18.43°
3. 63.44°
4. 54.74°
5. 6.143

Exercise 3.2

1. $2(x-1)+3(y-1)-(z-1)=0$
2. $1(x-1)-1(y-2)=0$
3. $\frac{1}{3}[1,1,1]$
4. $z=5$

Chapter Three: Vectors and Scalars

Introduction:

In this chapter we shall use the ideas of the plane to develop a new mathematical concept, vector. If you have studied physics, you have encountered this concept in that part of physics concerned with forces and equilibrium.

Physicists were responsible for first conceiving the idea of a vector, but the mathematical concept of vectors has become important in its own right and has extremely wide application, not only in the sciences but in mathematics as well.

Scalars and Vectors:

A quantity which is completely specified by a certain number associated with a suitable unit without any mention of direction in space is known as scalar. Examples of scalar are time, mass, length, volume, density, temperature, energy, distance, speed etc. The number describing the quantity of a particular scalar is known as its magnitude. The scalars are added subtracted, multiplied and divided by the usual arithmetical laws.

A quantity which is completely described only when both their magnitude and direction are specified is known as vector. Examples of vector are force, velocity, acceleration, displacement, torque, momentum, gravitational force, electric and magnetic intensities etc. A vector is represented by a Roman letter in bold face and its magnitude, by the same letter in italics. Thus **V** means vector and V is magnitude.

Vector Representations:

A vector quantity is represented by a straight line segment, say $\overrightarrow{PQ}$. The arrow head indicate the direction from P to Q. The length of the Vector represents its magnitude. Sometimes the vectors are represented by single letter such as V or $\vec{V}$. The magnitude of a vector is denoted by |V| or by just **V**, where | $\vec{V}$ | means modulus of $\vec{V}$ which is a positive value

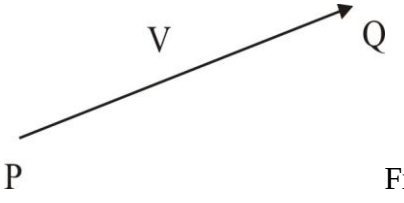


Fig. 1

Types of Vectors:

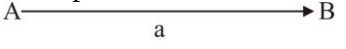
1. Unit Vector:

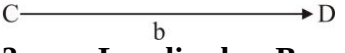
A vector whose magnitude is unity i.e., 1 and direction along the given vector is called a unit Vector. If $\overrightarrow{a}$ is a vector then a unit vector in the direction of $\overrightarrow{a}$, denoted by $\hat{a}$ (read as a cap), is given as,

$$\hat{a} = \frac{\overrightarrow{a}}{|a|} \quad \text{or} \quad \overrightarrow{a} = |a|\,\hat{a}$$

2. Free Vector:

A vector whose position is not fixed in space. Thus, the line of action of a free vector can be shifted parallel to itself. Displacement is an example of a free vector as shown in figure 1:





3. Localized or Bounded Vectors:

A vector which cannot be shifted parallel to itself, i.e., whose line of action is fixed is called a localized or bounded vector. Force and momentum are examples of localized vectors.

4. Coplanar Vectors:

The vectors which lies in the same plane are called coplanar vectors, as shown in Fig. 2.

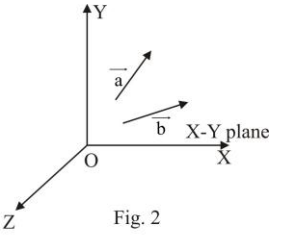


Fig. 2

5. Concurrent Vectors:

The vectors which pass through the common point are called concurrent vectors. In the figure no.3 vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ are called concurrent as they pass through the same point.

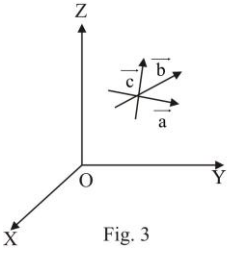


Fig. 3

6. Negative of a Vector:

The vector which has the same magnitude as the vector $\overrightarrow{a}$ but opposite in direction to $\overrightarrow{a}$ is called the negative to $\overrightarrow{a}$. It is represented by $-\overrightarrow{a}$. Thus of $\overrightarrow{AB} = -\overrightarrow{a}$ then $\overrightarrow{BA} = -\overrightarrow{a}$

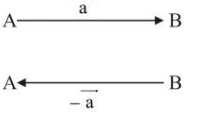


Fig. 4

7. Null or Zero Vector:

It is a vector whose magnitude is zero. We denote the null vector by O. The direction of a zero vector is arbitrary.

The vectors other than zero vectors are proper vectors or non-zero vectors.

8. Equal Vectors:

Two vectors $\vec{a}$ and $\vec{b}$ are said to be equal if they have the same magnitude and direction. If $\vec{a}$ and $\vec{b}$ are equal vectors then $\vec{a} = \vec{b}$

9. Parallel and Collinear Vectors:

The vectors $\vec{a}$ and $\vec{b}$ are parallel if for any real number n, $\vec{a} = n\vec{b}$. If

(i) $n > 0$ then the vectors $\vec{a}$ and $\vec{b}$ have the same direction.

(ii) $n < 0$ then $\vec{a}$ and $\vec{b}$ have opposite directions.

Now, we can also define collinear vectors which lie along the same straight line or having their directions parallel to one another.

10. Like and Unlike Vectors:

The vectors having same direction are called like vectors and those having opposite directions are called unlike vectors.

11. Position Vectors (PV):

If vector $\overrightarrow{OA}$ is used to specify the position of a point A relative to another point O. This $\overrightarrow{OA}$ is called the position vector of A referred to O as origin. In the figure 4 $\vec{a} = \overrightarrow{OA}$ and $\overrightarrow{OB} = \vec{b}$ are the position vector (P.V) of A and B respectively. The vector $\overrightarrow{AB}$ is determined as follows:

By the head and tail rules,

$\overrightarrow{OA} + \overrightarrow{AB} = \overrightarrow{OB}$

Or $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = \vec{b} - \vec{a}$

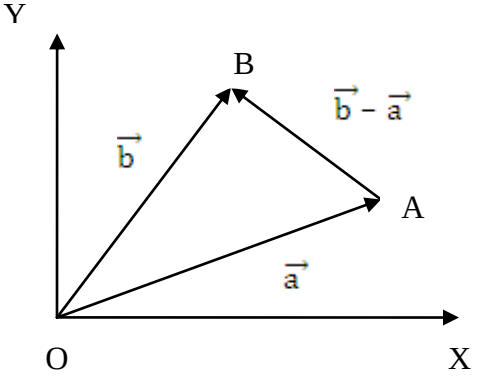


Fig. 5

Addition and Subtraction of Vectors:

1. Addition of Vectors:

Suppose $\vec{a}$ and $\vec{b}$ are any two vectors. Choose point A so that $\vec{a} = \overrightarrow{OA}$ and choose point C so that $\vec{b} = \overrightarrow{AC}$. The sum, $\vec{a} + \vec{b}$ of $\vec{a}$ and $\vec{b}$ is the vector is the vector $\overrightarrow{OC}$. Thus the sum of two vectors $\vec{a}$ and $\vec{b}$ is performed by the Triangle Law of addition.

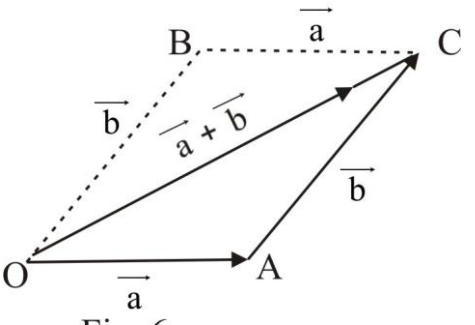


Fig. 6

2. Subtraction of Vectors:

If a vector $\vec{b}$ is to be subtracted from a vector $\vec{a}$, the difference vector $\vec{a} - \vec{b}$ can be obtained by adding vectors $\vec{a}$ and $-\vec{b}$.

The vector $-\vec{b}$ is a vector which is equal and parallel to that of vector but its arrow-head points in opposite direction. Now the vectors $\vec{a}$ and $-\vec{b}$ can be added by the head-to-tail rule. Thus the line $\overrightarrow{AC}$ represents, in magnitude and direction, the vector $\vec{a} - \vec{b}$.

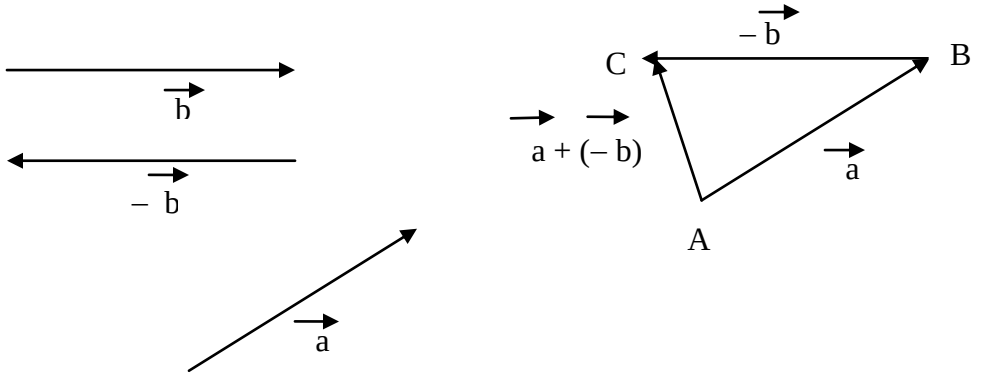


Fig . 7

Properties of Vector Addition:

i. Vector addition is commutative

i.e., $\vec{a} + \vec{b} = \vec{a} + \vec{b}$ where $\vec{a}$ and $\vec{b}$ are any two vectors.

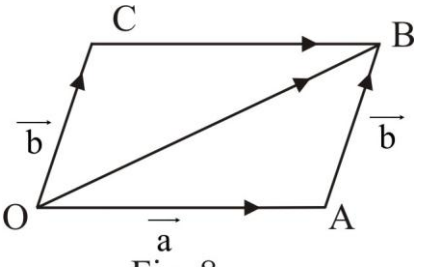


Fig. 8

(ii) Vectors Addition is Associative:

i.e. $\left(\vec{a} + \vec{b}\right) + \vec{c} = \vec{a} + \left(\vec{b} + \vec{c}\right)$

where $\vec{a}$, $\vec{b}$ and $\vec{c}$ are any three vectors.

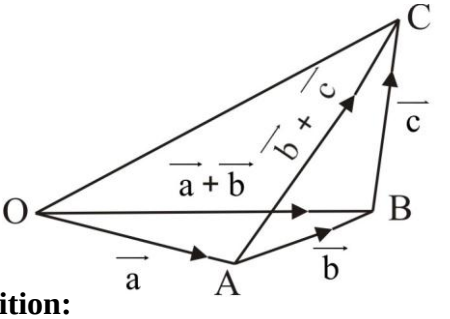


Fig.9

(iii) $\vec{O}$ is the identity in vectors addition:

For every vector $\vec{a}$

$\vec{a} + \vec{O} = \vec{a}$

Where $\vec{O}$ is the zero vector.

Remarks: Non-parallel vectors are not added or subtracted by the ordinary algebraic Laws because their resultant depends upon their directions as well.

6.6 Multiplication of a Vector by a Scalar:

If $\vec{a}$ is any vectors and K is a scalar, then $K\vec{a} = \vec{a}K$ is a vector with magnitude $|K|.|a|$ i.e., $|K|$ times the magnitude of $\vec{a}$ and whose direction is that of vector $\vec{a}$ or opposite to vector $\vec{a}$ according as K is positive or negative resp. In particular $\vec{a}$ and $-\vec{a}$ are opposite vectors.

Properties of Multiplication of Vectors by Scalars:

1. The scalar multiplication of a vectors satisfies

$$m(n\vec{a}) = (mn)\vec{a} = n(m\vec{a})$$

2. The scalar multiplication of a vector satisfies the distributive laws

i.e., $(m + n)\vec{a} = m\vec{a} + n\vec{a}$

and $m(\overrightarrow{a} + \overrightarrow{b}) = m\overrightarrow{a} + m\overrightarrow{b}$

Where m and n are scalars and $\overrightarrow{a}$ and $\overrightarrow{b}$ are vectors.

The Unit Vectors i, j, k (orthogonal system of unit Vectors):

Let us consider three mutually perpendicular straight lines OX, OY and OZ. These three mutually perpendicular lines determine uniquely the position of a point. Hence these lines may be taken as the co-ordinates axes with O as the origin.

We shall use i, j and k to denote the Unit Vectors along OX, OY and OX respectively.

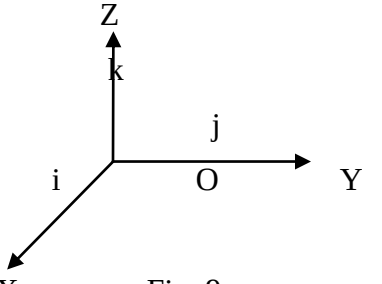


Fig. 9a

Representation of a Vector in the Form of Unit Vectors i, j and k.

Let us consider a vector $\overrightarrow{r} = \overrightarrow{OP}$ as shown in fig. 11. Then x i, y j and z k are vectors directed along the axes,

$$\overrightarrow{OQ} = \overrightarrow{OA} + \overrightarrow{AQ} = \overrightarrow{OA} + \overrightarrow{OB}$$ because

and $\overrightarrow{OQ} = xi + yi$

Because $\overrightarrow{QP} = zk$

$\overrightarrow{OP} = \overrightarrow{OQ} + \overrightarrow{QP}$

and $\overrightarrow{r} = \overrightarrow{OP} = xi + yj + zk$

Here the real numbers x, y and z are the components of Vector $\overrightarrow{r}$ or the co-ordinates of point P in the direction of OX, OY and OZ respectively. The vectors xi, yj and zk are called the resolved parts of the vector $\overrightarrow{r}$ in the direction of the Unit vectors i, j and k respectively.

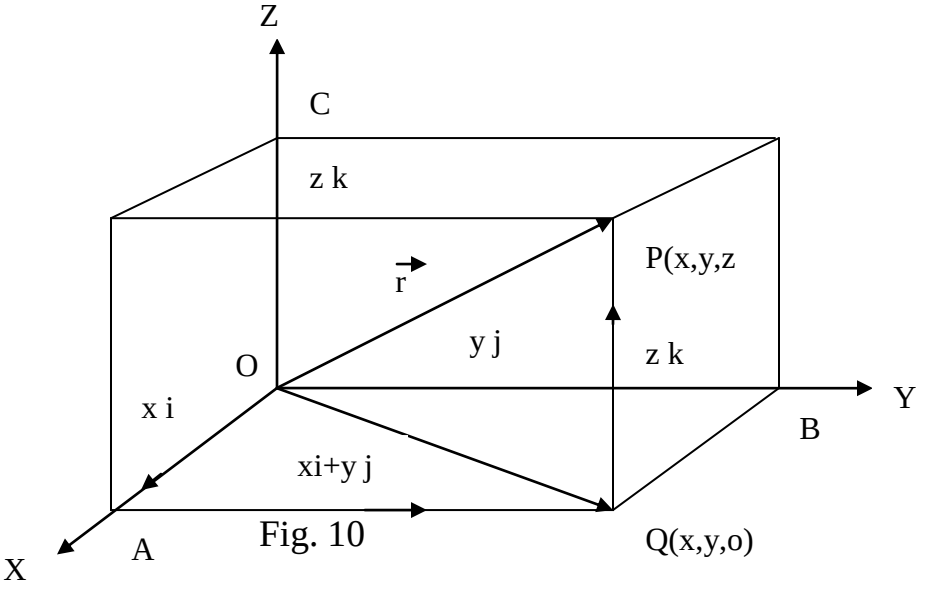


Fig. 10

6.9 Components of a Vector when the Tail is not at the Origin:

Consider a vector $\overrightarrow{r} = \overrightarrow{PQ}$ whose tail is at the point $P(x_1, y_1, z_1)$ and the head at the point $Q(x_2, y_2, z_2)$. Draw perpendiculars PP′ and QQ′ on x-aixs.

$P'Q' = x_2 - x_1$ = x-component of $\overrightarrow{r}$

Now draw perpendiculars PP^o and QQ^o on y-axis.

Then $P^oQ^o = y_2 - y_1$ = y-component of $\overrightarrow{r}$

Similarly $z_2 - z_1$ = z-component of $\overrightarrow{r}$

Hence the vector $\overrightarrow{r}$ can be written as,

$\overrightarrow{r} = \overrightarrow{PQ} = (x_2 - x_1)i + (y_2 - y_1)j + (z_2 - z_1)k$

Or, $\overrightarrow{r} = \overrightarrow{PQ} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$

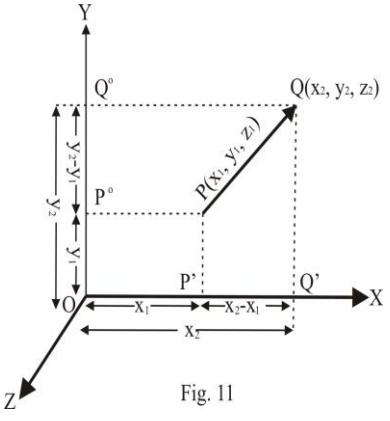


Fig. 11

6.10 Magnitude or Modulus of a Vector:

Suppose x, y and z are the magnitude of the vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$ as shown in fig. 10.

In the right triangle OAQ, by Pythagorean Theorem

$$OQ^2 = x^2 + y^2$$

Also in the right triangle OQP, we have

$$OP^2 = OQ^2 + QP^2$$

$$OP^2 = x^2 + y^2 + z^2$$

Or $$\left|\overrightarrow{r}\right| = |OP| = \sqrt{x^2 + y^2 + z^2}$$

Thus if $$\overrightarrow{r} = \overrightarrow{PQ} = xi + yj + zk$$

Then , its magnitude is

$$\left|\overrightarrow{r}\right| = \sqrt{x^2 + y^2 + z^2}$$

If $$\overrightarrow{r} = (x_2 - x_1)i + (y_2 - y_1)j + (z_2 - z_1)k$$

Then $$\left|\overrightarrow{r}\right| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Example 1:

If $P_1 = P(7, 4, -1)$ and $P_2 = P(3, -5, 4)$, what are the components of P_1P_2? Express $\overrightarrow{P_1\ P_2}$ in terms of i, j and k.

Solution:

x-component of $P_1\ P_2 = x_2 - x_1 = 3 - 7 = -4$

y-component of $P_1\ P_2 = y_2 - y_1 = -5 - 4 = -9$

and z-component of $P_1\ P_2 = z_2 - z_1 = 4 - (-1) = 5$

also $$\overrightarrow{P_1\ P_2} = (x_2 - x_1)i + (y_2 - y_1)j + (z_2 - z_1)k$$

$$\overrightarrow{P_1\ P_2} = -4i - 9j + 5k$$

Example 2: Find the magnitude of the vector

$$\left|\overrightarrow{u}\right| = \frac{3}{5}i - \frac{2}{5}j + \frac{2\sqrt{3}}{5}k$$

Solution:

$$\left|\overrightarrow{u}\right| = \sqrt{\left(\frac{3}{5}\right)^2 + \left(-\frac{2}{5}\right)^2 + \left(\frac{2\sqrt{3}}{5}\right)^2}$$

$$= \sqrt{\frac{9}{25} + \frac{4}{25} + \frac{12}{25}} = \sqrt{\frac{25}{25}}$$

$$\left|\overrightarrow{u}\right| = 1$$

Note: Two vectors are equal if and only if the corresponding components of these vectors are equal relative to the same co-ordinate system.

Example 3:

Find real numbers x, y and z such that

$xi + 2yj - zk + 3i - j = 4i + 3k$

Solution:

Since $(x + 3)i + (2y - 1)j + (-z)k = 4i + 3k$

Comparing both sides,we get

$x + 3 = 4,\ 2y - 1 = 0,\quad -z = 3$

$x = 1,\ y = \frac{1}{2},\ z = -3$

Note 2: If $\overrightarrow{r_1} = x_1 i + y_1 j + z_1 k$

$\overrightarrow{r_2} = x_2 i + y_2 j + z_2 k$

Then the sum vector =

$$\overrightarrow{r_1} + \overrightarrow{r_2} = (x_1 + x_2)i + (y_1 + y_2)j + (z_1 + z_2)k$$

Or $$\overrightarrow{r_1} + \overrightarrow{r_2} = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$$

Example 4:

$\overrightarrow{a} = 3i - 2j + 5k$ and $\overrightarrow{b} = -2i - j + k$.

Find $2\overrightarrow{a} - 3\overrightarrow{b}$ and also its unit vector.

Solution:

$$\begin{aligned} 2\overrightarrow{a} - 3\overrightarrow{b} &= 2(3i - 2j + 5k) - 3(-2i - j + k) \\ &= 6i - 4j + 10k + 6i + 3j - 3k \\ &= 12i - j + 7k \end{aligned}$$

If we denote $2\overrightarrow{a} - 3\overrightarrow{b} = \overrightarrow{c}$, then $\overrightarrow{c} = 12i - j + 7k$

and $|\vec{c}| = \sqrt{12^2 + (-1)^2 + 7^2} = \sqrt{144 + 1 + 49} = \sqrt{194}$

Therefore, $\hat{c} = \frac{\overrightarrow{c}}{|c|} = \frac{12i - j + 7k}{\sqrt{194}}$

$$\hat{c} = \frac{12}{\sqrt{194}}i - \frac{1}{\sqrt{194}}j + \frac{1}{\sqrt{194}}k$$

Note 3: Two vectors $\vec{r_1} = x_1 i + y_1 j + z_1 k$ and $\vec{r_2} = x_2 i + y_2 j + z_2 k$ are parallel if and only if $\frac{x_1}{x_2} = \frac{y_1}{y_2} = \frac{z_1}{z_2}$.

6.11 Direction Cosines:

Let us consider that the vector $\vec{r} = \overrightarrow{OP}$ which makes angles α, β and γ with the coordinate axes OX, OY and OZ respectively. Then Cos α, Cos β and Cos γ are called the direction cosines of the vector $\overrightarrow{OP}$. They are usually denoted by l, m and n respectively.

If $\overrightarrow{OP} = \vec{r} = xi + yj + zk$, then x, y and z are defined as the direction ratios of the vector $\vec{r}$ and $\vec{r} = \sqrt{x^2 + y^2 + z^2}$. Since the angles A, B and C are right angles (by the fig. 11), so in the right triangles.

OAP, OBP and OCP the direction cosines of $\vec{r}$ can be written as,

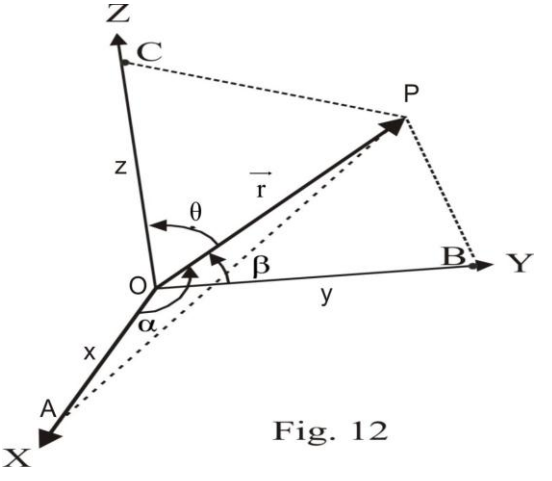


Fig. 12

$$l = \cos\alpha = \frac{x}{|r|} = \frac{x}{\sqrt{x^2 + y^2 + z^2}}$$

$$m = \cos\beta = \frac{y}{|r|} = \frac{y}{\sqrt{x^2 + y^2 + z^2}}$$

and $$n = \cos\gamma = \frac{z}{|r|} = \frac{z}{\sqrt{x^2 + y^2 + z^2}}$$

Note 1: Since the unit vector $\hat{r} = \frac{\vec{r}}{|r|} = \frac{xi + yj + zk}{|r|}$

$$\hat{r} = \frac{x}{|r|}i + \frac{y}{|r|}j + \frac{z}{|r|}k$$

$$\hat{r} = \text{Cos } \alpha \text{ i} + \text{Cos } \beta \text{ j} + \text{Cos } \gamma \text{ k}$$

Or $$\hat{r} = li + mj + nk$$

Therefore the co-efficient of i, j and k in the unit vector are the direction cosines of a vector.

Note 2: $$l^2 + m^2 + n^2 = \frac{x^2}{|r|^2} + \frac{y^2}{|r|^2} + \frac{z^2}{|r|^2}$$

$$= \frac{x^2 + y^2 + z^2}{|r|^2} = \frac{x^2 + y^2 + z^2}{x^2 + y^2 + z^2} = 1$$

Example 5:

Find the magnitude and direction cosines of the vectors $3i + 7j - 4k$, $i - 5j - 8k$ and $6i - 2j + 12k$.

Solution:

Let $\overrightarrow{a} = 3i + 7j - 4k$

$\overrightarrow{b} = i - 5j - 8k$

$\overrightarrow{c} = 6i - 2j + 12k$

Now $\hat{a} = \frac{\overrightarrow{a}}{|a|} = \frac{3i + 7j - 4k}{\sqrt{74}}$

$= \frac{3}{\sqrt{74}}i + \frac{7}{\sqrt{74}}j - \frac{4}{\sqrt{74}}k$

So the direction cosines of $\overrightarrow{a}$ are: $\frac{3}{\sqrt{74}}, \frac{7}{\sqrt{74}}, -\frac{4}{\sqrt{74}}$

Similarly the direction cosines of $\overrightarrow{b}$ are: $\frac{1}{\sqrt{90}}, -\frac{5}{\sqrt{90}}, -\frac{8}{\sqrt{90}}$

and the direction cosines of $\overrightarrow{c}$ are: $\frac{6}{\sqrt{184}}, -\frac{2}{\sqrt{184}}, \frac{12}{\sqrt{184}}$

Exercise 6.1

Q.1 If $\overrightarrow{a} = 3i - j - 4k$, $b = -2i + 4j - 3k$ and $\overrightarrow{c} = i + 2j - k$.
Find unit vector parallel to $3\overrightarrow{a} - 2\overrightarrow{b} + 4\overrightarrow{c}$.

Q.2 Find the vector whose magnitude is 5 and which is in the direction of the vector $4i - 3j + k$.

Q.3 For what value of m, the vector $4i + 2j - 3k$ and $mi - j + \sqrt{3}k$ have same magnitude?

Q.4 Given the points A = (1, 2, −1), B = (−3, 1, 2) and C = (0, −4, 3)
(i) find $\overrightarrow{AB}$, $\overrightarrow{BC}$, $\overrightarrow{AC}$ (ii) Show that $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$

Q.5 Find the lengths of the sides of a triangle, whose vertices are A = (2, 4, –1), B = (4, 5, 1), C = (3, 6, –3) and show that the triangle is right angled.

Q.6 If vectors $3i + j - k$ and $\lambda i - 4j + 4k$ are parallel, find the value of λ.

Q.7 Show that the vectors $4i - 6j + 9k$ and $-6i + 9j - \frac{27}{2}k$ are parallel.

Q.8 Find real numbers x, y and z such that

(a) $7xi + (y - 3)j + 6k = 10i + 8j - 3zk$

(b) $(x + 4)i + (y - 5)j + (z - 1)k = 0$

Q.9 Given the vectors $\vec{a} = 3i - 2j + 4k$ and $\vec{b} = 2i + j + 3k$ find the magnitude and direction cosines of

(i) $\vec{a} - \vec{b}$ (ii) $3\vec{a} - 2\vec{b}$

Q.10 If the position vector of $\vec{A}$ and $\vec{B}$ $5i - 2j + 4k$ and $i + 3j + 7k$ respectively, find the magnitude and direction cosines of $\overrightarrow{AB}$.

Answers 6.1

Q.1 $\frac{1}{\sqrt{398}}(17i - 3j - 10k)$ Q.2 $\frac{5}{\sqrt{26}}(4i - 3j + k)$

Q.3 ± 5 Q.4 $(-4, -1, 3), (3, -5, 1), (-1, -6, 4)$

Q.5 AB = AC = 3, BC = $3\sqrt{2}$ Q.6 $\lambda = -12$

Q.8 (a) $x = \frac{10}{7}$ y = 11 z = - 2 (b) x = -4 , y = 5 , z = 1

Q.9 (a) $\sqrt{11}; \frac{1}{\sqrt{11}} \frac{-3}{\sqrt{11}} \frac{1}{\sqrt{11}}$ (b) $5\sqrt{5}; \frac{1}{\sqrt{5}} \frac{-8}{5\sqrt{5}} \frac{6}{5\sqrt{5}}$

Q.10 $\sqrt{50}; \frac{-4}{\sqrt{50}}, \frac{5}{\sqrt{50}}, \frac{3}{\sqrt{50}}$

Product of Vectors:

1. Scalar Product of two Vectors:

If $\vec{a}$ and $\vec{b}$ are non-zero vectors, and θ is the angle between them, then the scalar product of $\vec{a}$ and $\vec{b}$ is denoted by $\vec{a} \cdot \vec{b}$ and read as $\vec{a}$ dot $\vec{b}$. It is defined by the relation

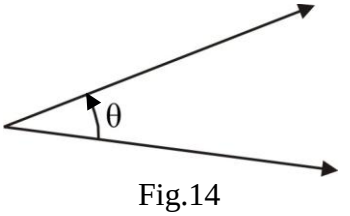


Fig.14

$$\vec{a} \cdot \vec{b} = |a|\,|b| \cos\theta \quad \ldots\ldots\ldots\ldots \quad (1)$$

If either $\vec{a}$ or $\vec{b}$ is the zero vector, then $\vec{a} \cdot \vec{b} = 0$

Remarks:

i. The scalar product of two vectors is also called the dot product because the "." used to indicate this kind of multiplication. Sometimes it is also called the inner product.

ii. The scalar product of two non-zero vectors is zero if and only if they are at right angles to each other. For $\vec{a} \cdot \vec{b} = 0$ implies that $\cos\theta = 0$, which is the condition of perpendicularity of two vectors.

Deductions:

From the definition (1) we deduct the following:

i. If $\vec{a}$ and $\vec{b}$ have the same direction, then

$\theta = 0^o \Rightarrow \cos 0^o = 1$

$\vec{a} \cdot \vec{b} = |\vec{a}|\,|\vec{b}|$

ii. If $\vec{a}$ and $\vec{b}$ have opposite directions, then

$\theta = \pi \Rightarrow \cos\pi = -1$

$\therefore \quad \vec{a} \cdot \vec{b} = -|\vec{a}|\,|\vec{b}|$

Iii $\vec{a} \cdot \vec{b}$ will be positive if $0 \le \theta < \frac{\pi}{2}$

and negative if, $\frac{\pi}{2} < \theta \le \pi$

iv The dot product of $\vec{a}$ and $\vec{b}$ is equal to the product of magnitude of $\vec{a}$ and the projection of $\vec{b}$ on $\vec{a}$. This illustrate the geometrical meaning of $\vec{a} \cdot \vec{b}$. In the fig. 15 $|\vec{b}|\cos\theta$ is the projection of $\vec{b}$ on $\vec{a}$.

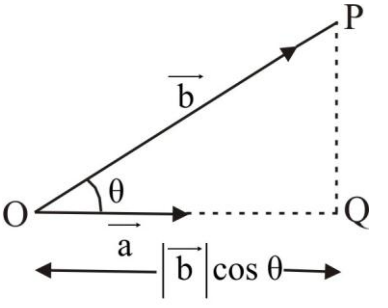


Fig. 15

v From the equation (1)

$\vec{b} \cdot \vec{a} = |\vec{b}|\,|\vec{a}|\cos\theta$

$= |\vec{a}|\,|\vec{b}|\cos\theta$

$\vec{b} \cdot \vec{a} = \vec{a} \cdot \vec{b}$

Hence the dot product is commutative.

Corollary 1:

If $\vec{a}$ be a vector, then the scalar product $\vec{a} \cdot \vec{a}$ can be expressed with the help of equation (1) as follows:

$$\vec{a} \cdot \vec{a} = \left|\vec{a}\right| \left|\vec{a}\right| \text{Cos } 0^o = \left|\vec{a}\right|^2$$

Or $$\left|\vec{a}\right| = = \sqrt{\vec{a} \cdot \vec{a}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

This relation gives us the magnitude of a vector in terms of dot product.

Corollary 2:

If i, j and k are the unit vectors in the directions of X−, Y− and Z− axes, then from eq. (2)

$$i^2 = i.i = |i| \ |i| \ \text{Cos}0^o$$

$$i^2 = 1$$

so $i^2 = j^2 = k^2 = 1$

and $i \cdot j = j \cdot i = 0$ Because $\text{Cos}90^o = 0$

$$i \cdot k = k \cdot i = 0$$

$$k \cdot i = i \cdot k = 0$$

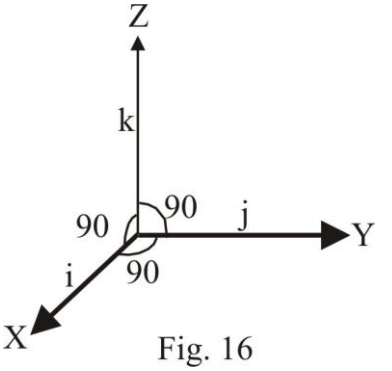


Fig. 16

Corollary 3:

(Analytical expression of $\vec{a} \cdot \vec{b}$)

Scalar product of two vectors in terms of their rectangular components.

For the two vectors

$$\vec{a} = a_1 i + a_2 j + a_3 k$$

and $$\vec{b} = b_1 i + b_2 j + b_3 k$$

the dot product is given as,

$$\vec{a} \cdot \vec{b} = (a_1 i + a_2 j + a_3 k) \cdot (b_1 i + b_2 j + b_3 k)$$

$$= a_1 b_1 + a_2 b_2 + a_3 b_3 \text{ as } i^2 = j^2 = k^2 = 1$$

and $i \cdot j = j \cdot k = k \cdot i = 0$

Also $\vec{a}$ and $\vec{b}$ are perpendicular if and only if $a_1 b_1 + a_2 b_2 + a_3 b_3 = 0$

Example 6:

If $\vec{a} = 3i + 4j - k$, $\vec{b} = -2i + 3j + k$ find $\vec{a} \cdot \vec{b}$

Solution:

$$\vec{a} \cdot \vec{b} = (3i + 4j - k) \cdot (-2i + 3j + k)$$

$$= -6 + 12 - 1$$

$$= 5$$

Example 7:

For what values of λ, the vectors $2i - j + 2k$ and $3i + 2\lambda j$ are perpendicular?

Solution:

Let $\vec{a} = 2i - j + 2k$ and $\vec{b} = 3i + 2\lambda j$

Since $\vec{a}$ and $\vec{b}$ are perpendicular,

So $\vec{a} \cdot \vec{b} = 0$

$(2i - j + 2k) \cdot (3i + 2\lambda j) = 0$

$6 - 2\lambda = 0$

Or $\lambda = 3$

Example 8:

Find the angle between the vectors $\vec{a}$ and $\vec{b}$, where $\vec{a} = i + 2j - k$ and $\vec{b} = -i + j - 2k$.

Solution:

As $\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos\theta$

Therefore $\cos\theta = \dfrac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|}$, $\vec{a} \cdot \vec{b} = -1 + 2 + 2 = 3$

$$|\vec{a}| = \sqrt{1+4+1} = \sqrt{6}, \quad |\vec{b}| = \sqrt{1+1+4} = \sqrt{6}$$

$$\cos\theta = \frac{3}{\sqrt{6}\sqrt{6}}$$

$$\cos\theta = \frac{3}{6} = \frac{1}{2}$$

$$\theta = \cos^{-1}\frac{1}{2} = 60^{o}$$

Example 9:

Consider the points A, B, C, D where coordinates are respectively (1, 1, 0), $(-1, 1, 0)$, $(1, -1, 0)$, $(0, -1, 1)$. Find the direction cosines of AC and BD and calculate the angle between them.

Solution:

Now we have A(1, 1, 0), B$(-1, 1, 0)$, C$(1, -1, 0)$, D$(0, -1, 1)$

$$\vec{a} = \overrightarrow{AC} = (1-1)i + (-1-1)j + (0-0)k = -2j$$

$\therefore$ Unit vector along AC $= \dfrac{\overrightarrow{AC}}{|AC|} = \dfrac{-2j}{2} = -j$

$\therefore$ The direction cosines of AC are $0, -1, 0$

Now

$$\vec{b} = \overrightarrow{BD} = (0+1)i + (-1-1)j + (1-0)k = i - 2j + k$$

Unit vector along BD $= \dfrac{\overrightarrow{BD}}{|BD|} = \dfrac{i-2j+k}{\sqrt{1+4+1}} = \dfrac{i-2j+k}{\sqrt{6}}$

The direction cosines of BD are:

$$\frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}}$$

Let, θ be the angle between AC and BD then:

$$\text{Cos } \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}|\,|\vec{b}|} = \frac{\overrightarrow{AC} \cdot \overrightarrow{BD}}{|\overrightarrow{AC}|\,|\overrightarrow{BD}|}$$

$$= \frac{(-2j).(i-2j+k)}{(2)\sqrt{6}}$$

$$= \frac{(-2)(-2)}{(2)\sqrt{6}} = \frac{2}{\sqrt{6}}$$

$$\theta = \text{Cos}^{-1}\left(\frac{2}{\sqrt{6}}\right)$$

Example 10:

Show that if $|a + b| = |a - b|$ then a and b are perpendicular.

Solution:

We have $|a + b| = |a - b|$

$\therefore$ $|a + b|^2 = |a - b|^2$ taking square.

$a^2 + b^2 + 2a.b = a^2 + b^2 - 2a.b$

$4a.b = 0$ or $a.b = 0$

Hence $\bar{a}$ and $\bar{b}$ are perpendicular.

2. Vector Product:

If $\vec{a}$ and $\vec{b}$ are non-zero vectors and θ is the angle between $\vec{a}$ and $\vec{b}$, then the vector product of $\vec{a}$ and $\vec{b}$, denoted by $\vec{a} \times \vec{b}$, is the vector $\vec{c}$ which is perpendicular to the plane determined by $\vec{a}$ and $\vec{b}$. It is defined by the relation,

$$\vec{c} = \vec{a} \times \vec{b} = \left(|\vec{a}||\vec{b}|\text{Sin } \theta\right) n$$

Where $\left|\vec{a}\right|\left|\vec{b}\right|\sin\theta$ is the magnitude of $\vec{c}$ and n is the Unit Vector in the direction of $\vec{c}$. The direction of $\vec{c}$ is determined by the right hand rule.

The vector product is also called the 'cross product' or 'Outer product' of the vectors.

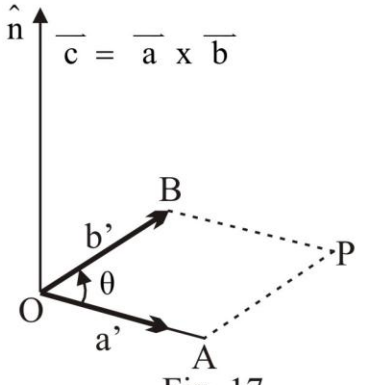


Fig. 17

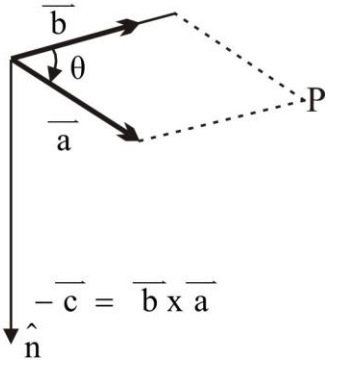


Fig. 18

Remarks:

If we consider $\vec{b} \times \vec{a}$, then $\vec{b} \times \vec{a}$ would be a vector which is opposite in the direction to $\vec{a} \times \vec{b}$.

Hence $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$

Which gives that $\vec{a} \times \vec{b} \neq \vec{b} \times \vec{a}$ in general

Hence the vector product is not commutative.

Deductions:

The following results may be derived from the definition.

i. The vector product of two non-zero vectors is zero if $\vec{a}$ and $\vec{b}$ are parallel, the angle between $\vec{a}$ and $\vec{b}$ is zero. Sin 0^o = 0, Hence $\vec{a} \times \vec{b} = 0$.

For a x b = 0 implies that $\sin\theta = 0$ which is the condition of parallelism of two vectors. In particular $\vec{a} \times \vec{a} = 0$. Hence for the unit vectors i, j and k,

i x i = j x j = k x k = 0

ii. If $\vec{a}$ and $\vec{b}$ are perpendicular vectors, then $\vec{a} \times \vec{b}$ is a vector whose magnitude is $\left|\vec{a}\right|\left|\vec{b}\right|$ and whose direction is such that the vectors a, b, a x b form a right-handed system of three mutually perpendicular

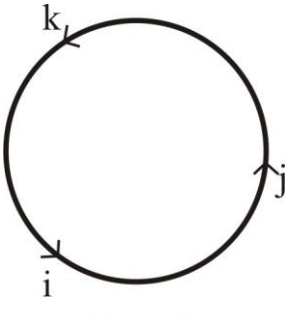


Fig. 19

vectors. In particular i x j = (1) (1) Sin 90^o k (k being perpendicular to i and j) = k

Similarly j x i = $-$k, i x k = $-$j, k x j = $-$i

Hence the cross product of two consecutive unit vectors is the third unit vector with the plus or minus sign according as the order of the product is anti-clockwise or clockwise respectively.

iii. Since $\left|\vec{a} \times \vec{b}\right| = \left|\vec{a}\right|\left|\vec{b}\right| \text{Sin}\,\theta$ …… (2)

Which is the area of the parallelogram whose two adjacent sides are $\left|\vec{a}\right|$ and $\left|\vec{b}\right|$.

Hence, area of parallelogram OABC = $\left|\vec{a} \times \vec{b}\right|$

and area of triangle OAB = $\frac{1}{2}\left|\vec{a} \times \vec{b}\right|$

If th vertices of a parallelogram are given, then

area of parallelogram OABC = $\left|\overrightarrow{OA} \times \overrightarrow{OB}\right|$

and, area of triangle OAB = $\frac{1}{2}\left|\overrightarrow{OA} \times \overrightarrow{OB}\right|$

iv. If n is the unit vector in the directions of $\vec{c} = \vec{a} \times \vec{b}$ then

$$n = \frac{\vec{c}}{|c|} = \frac{\vec{a} \times \vec{b}}{|a \times b|}$$

or $$n = \frac{\vec{a} \times \vec{b}}{|a||b|\text{Sin}\,\theta}$$

from equation (2) we also find.

$$\text{Sin}\,\theta = \frac{\left|\vec{a} \times \vec{b}\right|}{|a||b|}$$

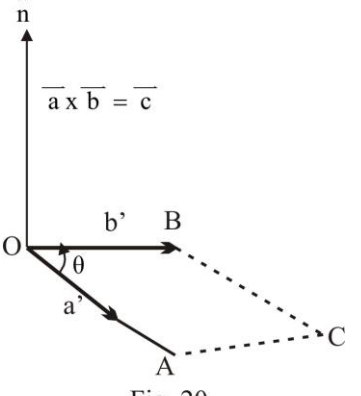


Fig. 20

6.13 Rectangular form of $\vec{a} \times \vec{b}$ (Analytical expression of $\vec{a} \times \vec{b}$)

If $\vec{a} = a_1i + a_2j + a_3k$

and $\vec{b} = b_1i + b_2j + b_3k$

then $\vec{a} \times \vec{b} = (a_1i + a_2j + a_3k) \times (b_1i + b_2j + b_3k)$

$= (a_1b_2k - a_1b_3j - a_2b_1k + a_2b_3j + a_3b_1j - a_3b_2j)$

$= (a_2b_3 - a_3b_2)i - (a_1b_3 - a_3b_1)j + (a_1b_2 - a_2b_1)k$

This result can be expressed in determinant form as

$$\vec{a} \times \vec{b} = \begin{vmatrix} i & j & k \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Example 11:

If $\vec{a} = 2i + 3j + 4k$ $\vec{b} = I - j + k$, Find

(i) $\vec{a} \times \vec{b}$

(ii) Sine of the angle between these vectors.

(iii) Unit vector perpendicular to each vector.

Solution:

(i) $$\vec{a} \times \vec{b} = \begin{vmatrix} i & j & k \\ 2 & 3 & 4 \\ 1 & -1 & 1 \end{vmatrix}$$

$$\vec{a} \times \vec{b} = i(3 + 4) - j(2 - 4) + k(-2 - 3)$$

$$= 7i + 2j - 5k$$

(ii) $$\text{Sin}\,\theta = \frac{\left|\vec{a} \times \vec{b}\right|}{\left|\vec{a}\right|\left|\vec{b}\right|} = \frac{\sqrt{7^2 + 2^2 + (-5)^2}}{\sqrt{2^2 + 3^2 + 4^2}.\sqrt{1^2 + (-1)^2 + 1^2}}$$

$$= \frac{\sqrt{78}}{\sqrt{29}\sqrt{3}}$$

$$\text{Sin}\,\theta = \sqrt{\frac{26}{29}}$$

(iii) If $\hat{n}$ is the unit vector perpendicular to $\vec{a}$ and $\vec{b}$ then

$$\hat{n} = \frac{\vec{a} \times \vec{b}}{\left|\vec{a} \times \vec{b}\right|} = \frac{7i + 2j - 5k}{\sqrt{78}}$$

Example 12:

$\vec{a} = 3i + 2k$, $\vec{b} = 4i + 4j - 2k$

$\vec{c} = i - 2j + 3k$, $\vec{d} = 2i - j + 5k$

Compute $(\vec{d} \times \vec{c}).(\vec{a} - \vec{b})$

Solution:

$$\vec{d} \times \vec{c} = \begin{vmatrix} i & j & k \\ 2 & -1 & 5 \\ 1 & -2 & 3 \end{vmatrix}$$

$$= i(-3 + 10) - j(6 - 5) + k(-4 + 1)$$

$$= 7i - j - 3k$$

Also $\overrightarrow{a} - \overrightarrow{b} = -i - 4j + 4k$

Hence $(\overrightarrow{d} \times \overrightarrow{c}).(\overrightarrow{a} - \overrightarrow{b}) = (7i - j - 3k).(-i - 4j + 4k)$

$= -7 + 4 - 12$

$= -15$

Example 13:

Find the area of the parallelogram with adjacent sides,

$\overrightarrow{a} = i - j + k$, and $\overrightarrow{b} = 2j - 3k$

Solution:

$$\overrightarrow{a} \times \overrightarrow{b} = \begin{vmatrix} i & j & k \\ 1 & -1 & 1 \\ 0 & 2 & -3 \end{vmatrix}$$

$= i(3 - 2) - j(-3 - 0) + k(2 + 0)$

$= i + 3j + 2k$

Area of parallelogram $= |\overrightarrow{a} \times \overrightarrow{b}| = \sqrt{1 + 9 + 4}$

$= \sqrt{14}$ square unit.

Example 14:

Find the area of the triangle whose vertices are A(0, 0, 0), B(1, 1, 1) and C(0, 2, 3)

Solution:

Since $\overrightarrow{AB} = (1 - 0, 1 - 0, 1 - 0)$

$= (1, 1, 1)$

and $\overrightarrow{AC} = (0 - 0, 2 - 0, 3 - 0)$

$\overrightarrow{AC} = (0, 2, 3)$

$$\overrightarrow{AB} \times \overrightarrow{AC} = \begin{vmatrix} i & j & k \\ 1 & 1 & 1 \\ 0 & 2 & 3 \end{vmatrix}$$

$= i(3 - 2) - j(3 - 0) + k(2 - 0)$

$= i - 3j + 2k$

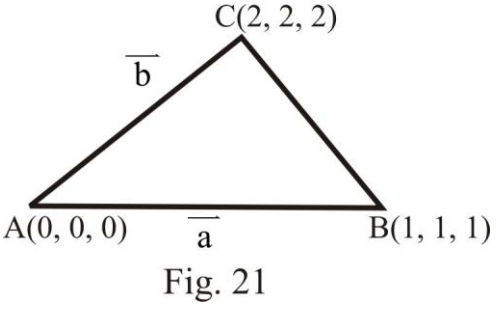


Fig. 21

Area of the triangle ABC $= \frac{1}{2} |\overrightarrow{AB} \times \overrightarrow{AC}| = \frac{1}{2}\sqrt{1^2 + (-3)^2 + 2^2}$

$= \frac{\sqrt{14}}{2}$ square unit

Example 15:

Prove by the use of cross-product that the points A(5, 2, −3), B(6, 1, 4), C(−2, −3, 6) and D(−3, −2, −1) are the vertices of a parallelogram.

Solution:

Since $\overrightarrow{AB} = (1, -1, 7)$

$\overrightarrow{DC} = (+1, -1, +7)$

$\overrightarrow{BC} = (-8, -4, 2)$

and $\overrightarrow{AD} = (-8, -4, 2)$

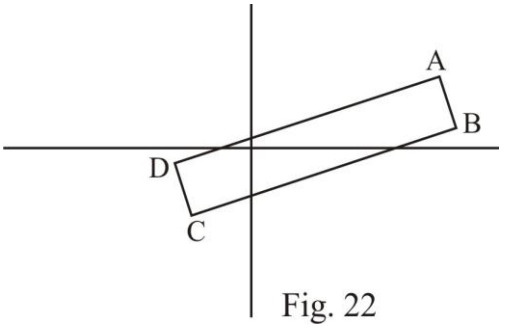


Fig. 22

$$\overrightarrow{AB} \times \overrightarrow{DC} = \begin{vmatrix} i & j & k \\ 1 & -1 & 7 \\ 1 & -1 & 7 \end{vmatrix}$$

$$= i(-7-7) - j(+7-7)+k(1-1)$$

$\overrightarrow{AB} \times \overrightarrow{DC} = 0$, so, $\overrightarrow{AB}$ and $\overrightarrow{DC}$ are parallel.

Also $$\overrightarrow{BC} \times \overrightarrow{AD} = \begin{vmatrix} i & j & k \\ -8 & -4 & 2 \\ -8 & -4 & 2 \end{vmatrix}$$

$$= i(0) - j(0)+k(0)$$

$\overrightarrow{BC} \times \overrightarrow{AD} = 0$, so, $\overrightarrow{BC}$ and $\overrightarrow{AD}$ are parallel.

Hence the given points are the vertices of a parallelogram.

Exercise 6.2

Q.1 Find $\vec{a} \cdot \vec{b}$ and $\vec{a} \times \vec{b}$

(i) $\vec{a} = 2i + 3j + 4k$ $\quad\quad$ $\vec{b} = i - j + k$

(ii) $\vec{a} = i + j + k$ $\quad\quad$ $\vec{b} = -5i + 2j - 3k$

(iii) $\vec{a} = -i - j - k$ $\quad\quad$ $\vec{b} = 2i + j$

Q.2 Show that the vectors $3i - j + 7k$ and $-6i + 3j + 3k$ are at right angle to each other.

Q.3 Find the cosine of the angle between the vectors:

(i) $\vec{a} = 2i - 8j + 3k$ $\quad\quad$ $\vec{b} = 4j + 3k$

(ii) $\vec{a} = i + 2j - k$ $\quad\quad$ $\vec{b} = -j - 2k$

(iii) $\vec{a} = 4i + 2j - k$ $\quad\quad$ $\vec{b} = 2i + 4j - k$

Q.4 If $\vec{a} = 3i + j - k$, $\vec{b} = 2i - j + k$ and $\vec{c} = 5i + 3k$, find $(2\vec{a} + \vec{b}) \cdot \vec{c}$.

Q.5 What is the cosine of the angle between $\overline{P_1P_2}$ and $\overline{P_3P_4}$ If $P_1(2,1,3)$, $P_2(-4, 4, 5)$, $P_3(0, 7, 0)$ and $P_4(-3, 4, -2)$?

Q.6 If $\vec{a} = [a_1, a_2, a_3]$ and $\vec{b} = [b_1, b_2, b_3]$, prove that:

$$\vec{a} \cdot \vec{b} = \frac{1}{2}\left[|\vec{a} + \vec{b}|^2 - |\vec{a}|^2 - |\vec{b}|^2\right]$$

Q.7 Find $(\vec{a} + \vec{b}).(\vec{a} - \vec{b})$ if $\vec{a} = i + 2j + 3k$ and $\vec{b} = 2i - j + k$.

Q.8 Prove that for every pair of vectors $\vec{a}$ and $\vec{b}$

$$(\vec{a} + \vec{b}).(\vec{a} - \vec{b}) = |\vec{a}|^2 - |\vec{b}|^2$$

Q.9 Find x so that $\vec{a}$ and $\vec{b}$ are perpendicular,

(i) $\vec{a} = 2i + 4j - 7k$ and $\vec{b} = 2i + 6j + xk$

(ii) $\vec{a} = xi - 2j + 5k$ and $\vec{b} = 2i - j + 3k$

Q.10 If $\vec{a} = 2i - 3j + 4k$ and $\vec{b} = 2j + 4k$

Find the component or projection of $\vec{a}$ along $\vec{b}$.

Q.11 Under what condition does the relation $(\vec{a} \cdot \vec{b})^2 = \vec{a}^2 \vec{b}^2$ hold for two vectors $\vec{a}$ and $\vec{b}$.

Q.12 If the vectors $3i + j - k$ and $\lambda I - 4 j + 4 k$ are parallel , find value of λ.

Q.13 If $\vec{a} = i - 2j + k$, $\vec{b} = i + 2j - 4k$, $\vec{c} = 2i - 3j + k$ Evaluate:

(i) $(\vec{a} \times \vec{b}).(\vec{a} \times \vec{c})$ (ii) $(\vec{a} \times \vec{b}) \times (\vec{a} \times \vec{c})$

Q.14 If $\vec{a} = i + 3j - 7k$ and $\vec{b} = 5i - 2j + 4k$. Find:

(i) $\vec{a} \cdot \vec{b}$ (ii) $\vec{a} \times \vec{b}$

(iii) Direction cosines of $\vec{a} \times \vec{b}$

Q.15 Prove that for the vectors $\vec{a}$ and $\vec{b}$

(i) $|\vec{a} \times \vec{b}|^2 + |\vec{a} \cdot \vec{b}|^2 = |\vec{a}|^2 |\vec{b}|^2$

(ii) $(\vec{a} - \vec{b}) \times (\vec{a} \times \vec{b}) = 2(\vec{a} \times \vec{b})$

Q.16 Prove that for vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$

$$[\vec{a} \times (\vec{b} + \vec{c})] + [\vec{b} \times (\vec{c} + \vec{a})] + [\vec{c} \times (\vec{a} + \vec{b})] = 0$$

Q.17 Find a vector perpendicular to both the lines AB and CD, where A is (0, 2, 4), B is (3, –1, 2), C is (2, 0, 1) and D is (4, 2, 0).

Q.18 Find $\left|(\vec{a} \times \vec{b}) \times \vec{c}\right|$ if $\vec{a} = i - 2j - 3k$, $\vec{b} = 2i + j - k$, $\vec{c} = i + 3j - 2k$.

Q.19 Find the sine of the angle and the unit vector perpendicular to each:

(i) $\vec{a} = i + j + k$ and $\vec{b} = 2i + 3j - k$

(ii) $\vec{a} = 2i - j + k$ and $\vec{b} = 3i + 4j - k$

Q.20 Given $\vec{a} = 2i - j$ and $\vec{b} = j + k$, if $\left|\vec{c}\right| = 12$ and $\vec{c}$ is perpendicular to both $\vec{a}$ and $\vec{b}$, write the component form of $\vec{c}$.

Q.21 Using cross product, find the area of each triangle whose vertices have the following co-ordinates:

(i) (0, 0, 0),(1, 1, 1),(0, 0, 3)

(ii) (2, 0, 0), (0, 2, 0), (0, 0, 2)

(iii) (1, -1, 1),(2, 2, 2), (4, -2, 1)

Q.22 Find the area of parallelogram determined by the vectors $\vec{a}$ and $\vec{b}$ $\vec{a} = i + 2j + 3k$ and $\vec{b} = -3i - 2j + k$.

Answers 6.2

Q.1 **(i)** 3; $7i - 2j - 5k$ **(ii)** $-6, -5i - 2i + 7k$ **(iii)** -3; $-i - 2j + k$

Q.3 **(i)** $-\frac{33}{5\sqrt{77}}$ **(ii)** 0 **(iii)** $\frac{17}{21}$ **Q.4** 37

Q.5 $\frac{5}{7\sqrt{22}}$ **Q.7** 8 **Q.9(i)**4 **(ii)** $-\frac{17}{2}$ **Q.10** $\sqrt{5}$

Q.11 [0, 11] **Q.12** $\lambda = -12$ **Q.13** (i) 15 (ii)$i - 2j + k$

Q.14 **(i)** –29 **(ii)** $-2i - 39j - 17k$ **(iii)** $-\frac{2}{\sqrt{1814}}, \frac{-39}{\sqrt{1814}}, \frac{-17}{\sqrt{1814}}$

Q.17 $7i - j + 12k$ **Q.18** $5\sqrt{26}$

Q.19 **(i)** $\sqrt{\frac{13}{21}}, \frac{-4i + 3j + k}{\sqrt{26}}$ **(ii)** $\sqrt{\frac{155}{156}}, \frac{-3i + 5j + 11k}{\sqrt{155}}$

Q.20 $-4i - 8j + 8k$

Q.21 **(i)** $\frac{3\sqrt{2}}{2}$ sq. unit. **(ii)** $2\sqrt{3}$ sq. unit.

(iii) $\frac{\sqrt{110}}{2}$ sq. unit. Q.22 $\sqrt{180}$ sq. unit

Summary

A vector is a quantity which has magnitude as well as direction while scalar is a quantity which has only magnitude. Vector is denoted as $\overrightarrow{AB}$ or $\overrightarrow{OP}$.

1. If P (x, y, z) be a point in space, then the position vector of P relative to 0 = $\overrightarrow{OP}$.
2. Unit coordinator vectors x, j, k are taken as unit vector s along axis $\overrightarrow{OP}$ = xi + yj + zk.
3. Magnitude of a vector. i.e. $|\overrightarrow{OP}| = \sqrt{x^2 + y^2 + z^2}$
4. Unit vector of a (non-zero vector), then $\hat{a} = \dfrac{\overrightarrow{a}}{|a|}$
5. Direction cosines of $\overrightarrow{OP}$ = xi + yj + zk then,

$$\cos\alpha = \frac{x}{\left|\overrightarrow{OP}\right|}. \text{Cos}\beta = \frac{y}{\left|\overrightarrow{OP}\right|}.\text{Cos}\gamma = \frac{z}{\left|\overrightarrow{OP}\right|}$$

Scalar product:

The scalar product of two vector $\overrightarrow{a}$ and $\overrightarrow{b}$ is defined as $\overrightarrow{a} . \overrightarrow{b} = |\overrightarrow{a}||\overrightarrow{b}| \text{Cos}\theta$

1. If a.b = 0, vectors are perpendicular.
2. i.j = j.k = i.k = zero while i.i = j.j = k.k = 1
3. $\overrightarrow{a} . \overrightarrow{b} = (a_1i + a_2j + a_3k).(b_1i + b_2j + b_3k) = a_1b_1 + a_2b_2 + a_3b_3$

Vector product:

The vector or cross product of two vectors $\overrightarrow{a}$ and $\overrightarrow{b}$ denoted $\overrightarrow{a} \text{ x } \overrightarrow{b}$ and is defined as: $\overrightarrow{a} \text{ x } \overrightarrow{b} = |\overrightarrow{a}||\overrightarrow{b}| \text{Sin}\theta \text{ n}$, $\text{Sin}\theta = \dfrac{|a \text{ x } b|}{|a||b|}$

1. $n = \dfrac{\overrightarrow{a} \text{ x } \overrightarrow{b}}{|a \text{ x } b|}$ unit vector.
2. $\overrightarrow{a} \text{ x } \overrightarrow{b} = 0$. $\overrightarrow{a}$ and $\overrightarrow{b}$ are parallel or collinear.
3. i x j = j x j = k x k = 0 and i x j = k.j x k = i, k x i = j
4. $\overrightarrow{a} \text{ x } \overrightarrow{b} = -\overrightarrow{b} \text{ x } \overrightarrow{a}$
5. $\overrightarrow{a} \text{ x } \overrightarrow{b} = (a_1i + a_2j + a_3k) \text{ x } (b_1i + b_2j + b_3k)$

$$= \begin{vmatrix} i & j & k \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Short Questions

Write the short answers of the following:

Q.1: What is scalar? Give examples.

Q.2: What is a vector? Give example.

Q.3: What is unit vector?

Q.4: Find the formula for magnitude of the vector $\vec{r} = xi + yj + zk$

Q.5: Find the magnitude of vector $-2i - 4j + 3k$

Q.6: What are parallel vectors?

Q.7: Find α, so that $|\alpha i + (\alpha + 1)j + 2k| = 3$

Q.8: If $\cos\alpha$, $\cos\beta$, $\cos\gamma$ are direction cosines of a vector

$\vec{r} = xi + yj + zk$, then show that $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$

Q.9: Find the unit vector along vector $4i - 3j - 5k$.

Q.10: Find the unit vector parallel to the sum of the vectors

$\bar{a} = [2, 4, -5]$, $\vec{b} = [1, 2, 3]$

Q.11: Given the vectors, $\bar{a} = 3i - 2j + k$, $\bar{b} = 2i - 4j - 3k$

$\bar{c} = -i + 2j + 2k$, Find $a + b + c$

Q.12: Given the vectors $\bar{a} = 3i + j - k$ and $\bar{b} = 2i + j - k$,find magnitude of $3\bar{a} - \bar{b}$

Q.13: Find a vector whose magnitude is 2 and is parallel to $5i + 3j + 2k$.

Q.14: Define scalar product of two vectors.

Q.15 Find $\bar{a} \cdot \bar{b}$ if $\bar{a} = i + 2j + 2k$, $\bar{b} = 3i - 2j - 4k$

Q.16: Find $(a + b) \cdot (a - b)$ if

$\bar{a} = 2i + 2j + 3k$, $\bar{b} = 2i - j + k$

Q.17: Define Vector product.

Q.18: If $\bar{a} = 2i + 3j + 4k$, $\bar{b} = i - j + k$

Find $|\vec{a} \times \vec{b}|$

Q.19: Find the area of parallelogram with adjacent sides,

$\bar{a}$ = 7i – j +k and $\bar{b}$ =2j – 3k

Q.20: For what value of λ, the vectors 2i – j + 2k and 3i +2 λj are perpendicular.

Q.21: Under what conditions does the relation $\bar{a} \cdot \bar{b} = |\bar{a}||\bar{b}|$ hold?

Q.22: Find scalars x and y such that x (i +2j)+ y(3i +4j) = 7

Q.23: Prove that if $\bar{a}$ = i + 3j – 2k and $\bar{b}$ = *i* – j –k, then $\bar{a}$ and $\bar{b}$ are perpendicular to each other.

Answers

5. $\sqrt{29}$ 7. 1 , – 2 9. $\frac{4i - 3j - 5k}{5\sqrt{2}}$

10. $\frac{3i + 6j - 2k}{7}$ 11. 4i – 4j + 0k 12. $\sqrt{54}$

13. $\frac{10\,i + 6j + 4k}{\sqrt{38}}$ 15. –9 16. 8

18. $\sqrt{78}$ 19. $\sqrt{14}$ sq. unit 20. λ = 3

21 θ = 0^o 22. x = $\frac{-1}{2}$, y =5/2

Objective Type Question

Q.1 Each questions has four possible answers. Choose the correct answer and encircle it.

__1. Magnitude of the vector $2i - 2j - k$ is:
(a) 4 (b) 3 (c) 2 (d) 1

__2. Unit vector of $i + j + k$ is:
(a) $i + j + k$ (b) $\frac{1}{3}(i + j + k)$
(c) $\frac{1}{\sqrt{3}}(i + j + k)$ (d) $\frac{1}{2}(i + j + k)$

__3. Unit vector of $i - 2j - 2k$ is:
(a) $i - 2j - 2k$ (b) $\frac{1}{3}(i - 2j - 2k)$
(c) $\frac{1}{\sqrt{3}}(i - 2j - 2k)$ (d) $\frac{1}{2}(i - 2j - 2k)$

__4. If $\hat{i}$, $\hat{j}$ and k are orthogonal unit vectors, then $j \times i$ is:
(a) k (b) $-k$ (c) 1 (d) -1

__5. The magnitude of a vector $\hat{i} - 3\hat{j} + 5k$ is:
(a) 3 (b) 25 (c) 35 (d) $\sqrt{35}$

__6. In l, m and n are direction cosine of a vector, then:
(a) $l^2 - m^2 - n^2 = 1$ (b) $l^2 - m^2 + n^2 = 1$
(c) $l^2 + m^2 - n^2 = 1$ (d) $l^2 + m^2 + n^2 = 1$

__7. If θ is the angle between the vector $\vec{a}$ and $\vec{b}$, then $\cos\theta$ is:
(a) $\vec{a} \cdot \vec{b}$ (b) $\frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|}$
(c) $\frac{\vec{a} \cdot \vec{b}}{|\vec{a}|}$ (d) $\frac{\vec{a} \cdot \vec{b}}{|\vec{b}|}$

__8. If $\vec{a} = a_1j + a_2j + a_3k$, $\vec{b} = b_1i + b_2j + b_3k$, then $\vec{a} \cdot \vec{b}$ is:
(a) $a_1b_1j + a_2b_2j + a_3b_3k$ (b) $a_1b_1 + a_2b_2 + a_3b_3$
(c) $a_1b_2j + a_2b_3j + a_3b_1k$ (d) None of these

__9. $\vec{a} \cdot \vec{b} = 0$ implies that $\vec{a}$ and $\vec{b}$ are:
(a) Perpendicular (b) Parallel

(c) Non-parallel (d) Oblique

__10. If $\vec{a} = i + j + k$ and $\vec{b} = -i - j - mk$ are perndicular then m will be equal to:

(a) 1 (b) -2 (c) ± 1 (d) ± 3

__11. $\vec{a} \cdot \vec{b}$ is a:

(a) Vector quantity (b) Scalar quantity
(c) Unity (d) None of these

__12. $\vec{a} \cdot \vec{a}$ is equal to:

(a) 1 (b) a^2 (c) $\left|\vec{a}\right|$ (d) None of these

__13. If $\vec{a} = 2i - 3j + k$ and $\vec{b} = -i + 2j + 7k$ then $\vec{a} \cdot \vec{b}$ is equal to:

(a) -1 (b) -2 (c) -3 (d) -4

__14. If $\vec{a} \times \vec{b} = 0$ then $\vec{a}$ and $\vec{b}$ is:

(a) Non-parallel (b) Parallel
(c) Perpendicular (d) None of these

__15. The cross product of two vectors $\vec{a}$ and $\vec{b}$ is:

(a) $\left|\vec{a}\right|\left|\vec{b}\right| \cos\theta$ (b) $\left|\vec{a}\right|\left|\vec{b}\right| \sin\theta$

(c) $\left|\vec{a}\right|\left|\vec{b}\right| \sin\theta\, \hat{n}$ (d) $\left|\vec{a}\right|\left|\vec{b}\right| \cos\theta\, \hat{n}$

__16. If $\hat{n}$ is the unit vector in the direction of $\vec{a} \times \vec{b}$, then $\hat{n}$ is:

(a) $\dfrac{\vec{a} \times \vec{b}}{\left|\vec{a}\right|\left|\vec{b}\right|}$ (b) $\dfrac{\vec{a} \cdot \vec{b}}{\left|\vec{a}\right|\left|\vec{b}\right|}$ (c) $\dfrac{\left|\vec{a} \times \vec{b}\right|}{\left|\vec{a}\right|\left|\vec{b}\right| \sin\theta}$ (d) $\dfrac{\vec{a} \times \vec{b}}{\left|\vec{a} \times \vec{b}\right|}$

__17. $\left|\vec{a} \times \vec{b}\right|$ is area of the figure called:

(a) Triangle (b) Rectangle (c) Parallelogram (d) Sector

__18. $\vec{a} \times \vec{b}$ is equal to:

(a) $-\vec{b} \times \vec{a}$ (b) $\vec{b} \times \vec{a}$ (c) $\left|\vec{a} \times \vec{b}\right|$ (d) $\left|\vec{b} \times \vec{a}\right|$

__19. If $\vec{a}$ and $\vec{b}$ are collinear vectors, then:

(a) $\vec{a} \times \vec{b} = 0$ (b) $\vec{a} \cdot \vec{b} = 0$
(c) $\vec{a} - \vec{b} = 0$ (d) $\vec{a} + \vec{b} = 0$

__20. $\left|\vec{a} \times \vec{b}\right|$ is a:

(a) Vector quantity (b) Scalar quantity
(c) Unity (d) None of these

Answers

1.	b	2.	c	3.	b	4.	b	5.	d
6.	d	7.	b	8.	b	9.	a	10.	c
11.	b	12.	b	13.	a	14.	b	15.	c
16.	d	17.	c	18.	a	19.	a	20.	b

Chapter Four:
Vectors and Translation Maps

Student Outcomes

- Students use vectors to define a translation map and translate geometric figures in $\mathbb{R}^2$ and $\mathbb{R}^3$.
- Students represent and perform calculations with three-dimensional vectors and understand operations in three dimensions that are analogous to operations in two dimensions.

Lesson Notes

This lesson builds on the use of vectors to represent shifts by using vectors to define translation maps studied in prior lessons. The connection between matrices and vectors should become more apparent in this lesson. A notation for vectors that recalls matrix notation is used. This lesson focuses on several N-VM standards including **N-VM.A.1**, **N-VM.A.3**, **N-VM.B.4a**, **N-VM.V.4c**, and **N-VM.B.5**, extending vector addition, subtraction, and scalar multiplication to $\mathbb{R}^3$. This lesson lends itself to working in a program capable of creating three-dimensional graphs, such as GeoGebra (v5.0 or later). Consider modeling these lessons on one computer or, if access to a lab is available, have students work on their own computers in the latter portion of this lesson.

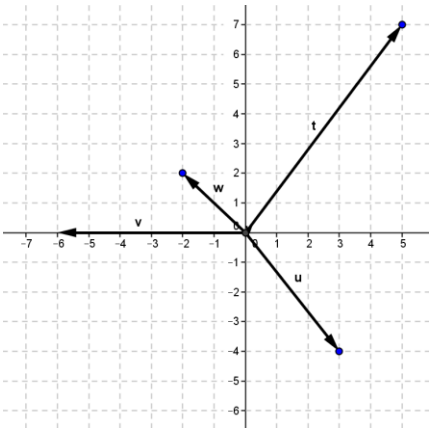
Classwork

Opening Exercise (5 minutes)

Activate student knowledge about the previous day's lesson with this short Opening Exercise. Students should work independently at first and then check their solutions with a partner. Students could draw their vectors using GeoGebra. If using GeoGebra, type vector [(**a**,**b**)] in the input box where **a** and **b** are the horizontal and vertical components of the vector. In GeoGebra, the vector $\mathbf{v} = \langle 1,2 \rangle$ appears as $\mathbf{v} = \binom{1}{2}$.

Opening Exercise

Write each vector described below in component form and find its magnitude. Draw an arrow originating from $(0, 0)$ to represent each vector's magnitude and direction.



a. **Translate 3 units right and 4 units down.**

$\mathbf{u} = \langle 3, -4 \rangle, \|\mathbf{u}\| = 5$

b. **Translate 6 units left.**

$\mathbf{v} = \langle -6, 0 \rangle, \|\mathbf{v}\| = 6$

c. **Translate 2 units left and 2 units up.**

$\mathbf{w} = \langle -2, 2 \rangle, \|\mathbf{w}\| = 2\sqrt{2}$

d. **Translate 5 units right and 7 units up.**

$\mathbf{t} = \langle 5, 7 \rangle, \|\mathbf{t}\| = \sqrt{25 + 49} = \sqrt{74}$

Discussion (5 minutes)

Lead a short discussion to model how to write a vector as a translation map. Introduce the column notation for vectors, and ask students to recall when they have seen something similar in previous lessons. They should recall it from Module 1 and from the beginning of this module.

- We can use a translation map to represent a vector. In fact, we can make this the definition of a vector: A vector is a translation map.
- For example, the vector $\mathbf{v} = \langle 2,1 \rangle$ is identified with the translation map, $T_{\mathbf{v}}$ that maps a point $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ to the point $\boldsymbol{x} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} x+2 \\ y+1 \end{bmatrix}$; thus

$$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+2 \\ y+1 \end{bmatrix}.$$

- Why would it make sense to think of a vector as a translation map?
  - *Because vectors accomplish the same thing as a translation map. An object is translated by the components of the vector in the horizontal and vertical direction. The object's size and shape would remain unchanged under this type of transformation.*
- Explain how writing vectors in a column makes it easier to think of them as translation maps.
  - *We are using the same kind of notation we used when working with translation maps and linear transformations in the previous module and earlier in this module.*

Give students time to summarize this information in their notes or on the student materials.

Scaffolding:

- GeoGebra or other graphing software can be a powerful tool for scaffolding this lesson, particularly for students who are struggling to quickly sketch graphs of circles and lines by hand.
- Post the formulas for a line and a circle on the board prior to this lesson, and have students pair share with a partner what they recall about these equations and the related graphs.
- A line with slope $-\frac{a}{b}$ and y-intercept $\left(0, \frac{c}{b}\right)$ is given by $ax + by = c$.
- A circle with radius r and center (h, k) is given by $(x-h)^2 + (y-k)^2 = r^2$.
- For the following equations, have students sketch the graph and state the key features.

$$2x - 6y = 12$$
$$x + 2y = -5$$
$$(x-1)^2 + (y-3)^2 = 9$$

Exercises 1–3 (5 minutes)

Provide students with the opportunity to work in small groups on the next exercises. Check for precision and accuracy in their work with the translation map notation applied to vectors. Make sure they are thinking carefully about the transformations of the circle under the mapping in Exercises 2 and 3, and check to make sure they are creating the equations of the image figures correctly.

Exercises 1–3

1. Write a translation map defined by each vector from the opening.

Consider the vector $\mathbf{v} = \langle -2, 5 \rangle$ and its associated translation map:

$$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - 2 \\ y + 5 \end{bmatrix}$$

Vector in Component Form	*Translation Map, $T_{\mathbf{v}}$*
$\mathbf{u} = \langle 3, -4 \rangle$	$T_{\mathbf{u}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + 3 \\ y - 4 \end{bmatrix}$
$\mathbf{v} = \langle -6, 0 \rangle$	$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - 6 \\ y \end{bmatrix}$
$\mathbf{w} = \langle -2, 2 \rangle$	$T_{\mathbf{w}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - 2 \\ y + 2 \end{bmatrix}$
$\mathbf{t} = \langle 5, 7 \rangle$	$T_{\mathbf{t}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + 5 \\ y + 7 \end{bmatrix}$

2. Suppose we apply the translation map $T_{\mathbf{v}}$ to each point on the circle $(x + 4)^2 + (y - 3)^2 = 25$.

a. What is the radius and center of the original circle?

The radius is 5 and the center is $(-4, 3)$.

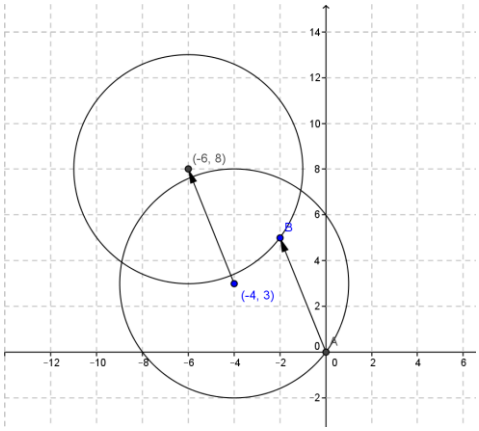
b. Show that the image points satisfy the equation of another circle.

Suppose that $T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x' \\ y' \end{bmatrix}$. Then $x' = x - 2$ and $y' = y + 5$, so $x = x' + 2$ and $y = y' - 5$.
Since $(x + 4)^2 + (y - 3)^2 = 25$, we have

$$(x + 4)^2 + (y - 3)^2 = 25$$
$$((x' + 2) + 4)^2 + ((y' - 5) - 3)^2 = 25$$
$$(x' + 6)^2 + (y' - 8)^2 = 25$$

So, the image points $\begin{bmatrix} x' \\ y' \end{bmatrix}$ lie on a circle.

c. What are the center and radius of this image circle?



The radius of 5 remains unchanged. The new center is $(-6, 8)$.

3. **Suppose we apply the translation map T_v to each point on the line $2x - 3y = 10$.**

   a. **What are the slope and y-intercept of the original line?**

   The slope is $\frac{2}{3}$ and the y-intercept is $\left(0, -\frac{10}{3}\right)$.

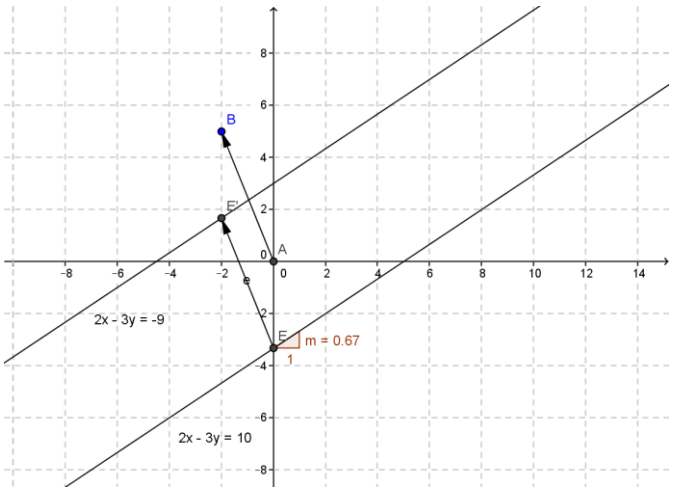
   b. **Show that the image points satisfy the equation of another line.**

   Suppose that $T_v\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x' \\ y' \end{bmatrix}$. Then $x' = x - 2$ and $y' = y + 5$, so $x = x' + 2$ and $y = y' - 5$.

   Since $2x - 3y = 10$, we have

$$2(x' + 2) - 3(y' - 5) = 10$$
$$2x' + 4 - 3y' + 15 = 10$$
$$2x' - 3y' = -9$$

   c. **What are the slope and y-intercept of this image line?**



   The slope is $\frac{2}{3}$ and the y-intercept is $(0, 3)$.

Discussion (5 minutes): Vectors in Three Dimensions

A vector in two dimensions can be used to define a translation map that translates objects in the coordinate plane by the horizontal and vertical components of the vector.

- What types of objects could we translate by a vector in two dimensions?
  - *Any geometric figure could be translated by a vector.*

This idea is easily extended to $\mathbb{R}^3$, the Cartesian coordinate system in three dimensions. For example, the vector $\mathbf{v} = \langle 1,3,5 \rangle$ is a translation in space 1 unit in the x direction, 3 units in the y direction and 5 units in the z direction.

- What would be the associated translation map for the vector $\mathbf{v} = \langle 1,3,5 \rangle$?
  - *The associated translation map would be* $\boldsymbol{T}_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y+3 \\ z+5 \end{bmatrix}$.
- What types of objects could we represent in three dimensions?
  - *We could represent points, planes, spheres, cubes, cylinders, pyramids, or other solid figures in three dimensions. Two-dimensional figures can also be represented in three dimensions.*

Give students an opportunity to summarize this information in their notes or on the student pages before starting the exercises that follow.

Example 1 (8 minutes): Vectors and Translation Maps in $\mathbb{R}^3$

If technology is available, model this example using GeoGebra by selecting *3-D Graphs* from the view menu. Type each figure into the input bar and then type in vector[(1,3,5)]. Use the *translate by vector* feature in the translate menu. Students will be able to see the equations updating and the image graph displayed along with the pre-image when using this feature. If technology is not available, use 3-dimensional graph paper or isometric graph paper. 3-D graph paper is included in Lesson 5 or can be downloaded for free at http://www.waterproofpaper.com/graph-paper/isometric-graphing-paper.pdf.

Scaffolding:

Help students graph in 3-D by using 3-D or isometric graph paper. 3-D graph paper is included in Lesson 5.

Take time to help students make sense of the graphs they are seeing of each object. Point out the similarities between points, planes, and spheres in three dimensions and points, lines, and circles in two dimensions.

- How do the coordinates of a point in two dimensions compare to coordinates in three dimensions?
  - *You just add another coordinate for the distance from the origin in the z direction.*

How does the equation of a line in two dimensions compare to the equation of a plane in three dimensions?

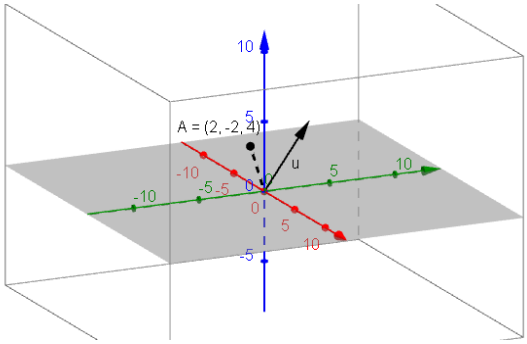
  - *You just add another linear term using the variable z. The intercepts are similar with two coordinates being 0 and the third coordinate giving the point where the plane crosses the given axis.*
- How does the equation of a circle in two dimensions compare to the equation of a sphere in three dimensions?
  - *You just add another quadratic term using the variable z. The radius is the cube root of the constant and the center is the same except three coordinates.*

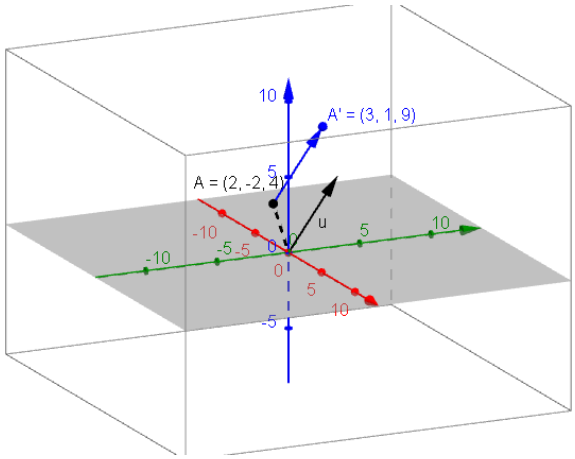
Example 1: Vectors and Translation Maps in $\mathbb{R}^3$

Translate by the vector $\mathbf{v} = \langle 1, 3, 5 \rangle$ by applying the translation map $T_{\mathbf{v}}$ to the following objects in $\mathbb{R}^3$. A sketch of the original object and the vector is shown. Sketch the image.

$$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y+3 \\ z+5 \end{bmatrix}$$

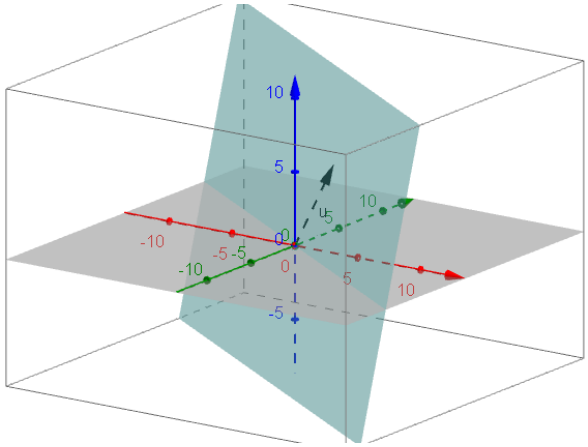
a. The point $A(2, -2, 4)$

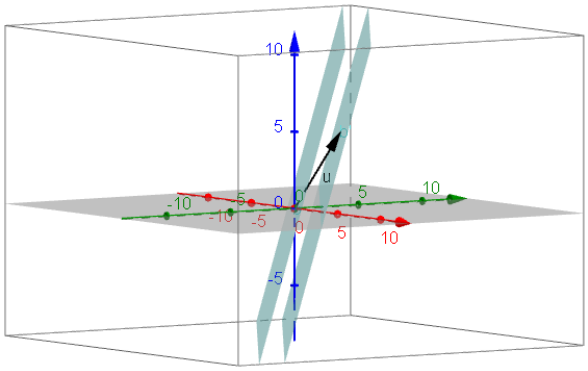




The new point will be $(2+1, -2+3, 4+5)$ or $(3, 1, 9)$.

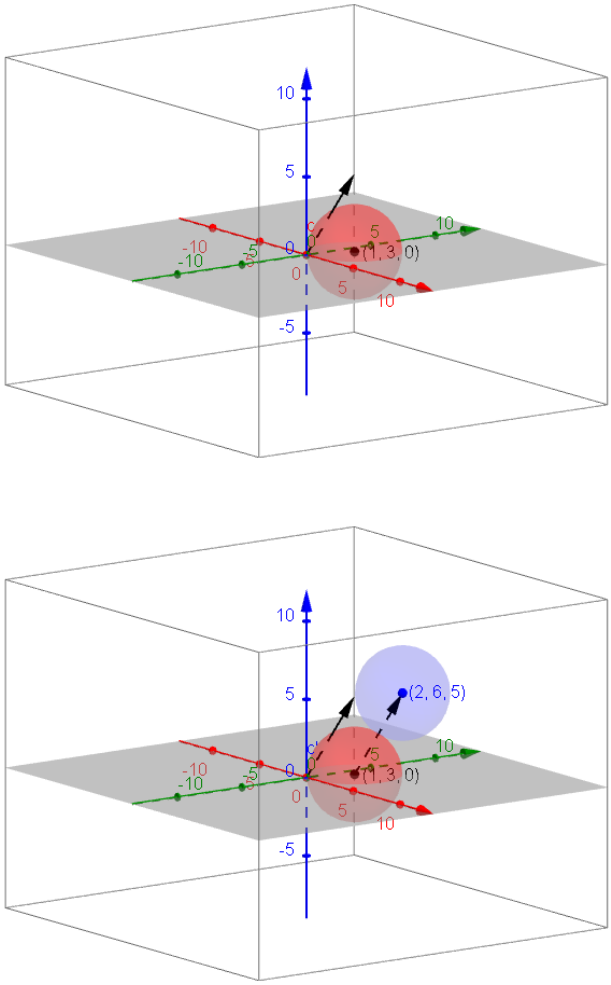
b. The plane $2x + 3y - z = 0$





The new plane will be $2(x-1)+3(y-3)-(z-5)=0$, which is equivalent to $2x+3y-z=6$.

c. **The sphere $(x-1)^2+(y-3)^2+z^2=9$**



The new sphere will be $\big((x-1)-1\big)^2+\big((y-3)-3\big)^2+(z-5)^2=9$***, which is equivalent to*** $(x-2)^2+(y-6)^2+(z-5)^2=9$.

After sharing the solutions to Example 1, conclude with a brief discussion of the results.

- When we translated a line by a vector, why was the image parallel to the original line?
  - *All points in the plane were moved by the same amount in the same direction due to the translation map.*
- Why does it make sense that, when we translate a plane by a vector, the image plane is parallel to the pre-image?
  - *If every point moves the same distance in the same direction, then the planes will have the same orientation in space but a different location.*

- When translating any geometric figure by a vector, will the image be congruent to the pre-image? Explain how we know.
  - *Since translations are rigid transformations, the image will be congruent to the pre-image.*

Exercise 4 (2 minutes)

This short exercise should reveal whether or not students are able to extend their thinking about vectors to three dimensions.

Exercise 4

4. Given the sphere $(x+3)^2+(y-1)^2+(z-3)^2=10$.

a. What are its center and radius?

The center is $(-3,1,3)$, and the radius is $\sqrt{10}$.

b. Write a vector and its associated translation map that would take this sphere to its image centered at the origin.

We need to translate the center from the point $(-3,1,3)$ to the point $(0,0,0)$. This represents a translation of 3 units in the x direction, -1 unit in the y direction, and -3 units in the z direction.

The vector is $\mathbf{v}=\langle 3,-1,-3\rangle$, and the translation map would be

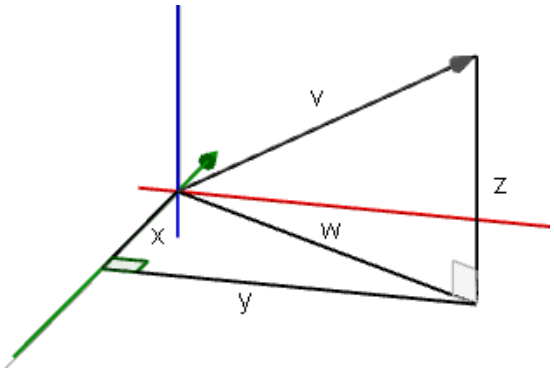
$$T_{\mathbf{v}}\left(\begin{bmatrix}x\\y\\z\end{bmatrix}\right)=\begin{bmatrix}x+3\\y-1\\z-3\end{bmatrix}.$$

Example 2 (5 minutes): What is the Magnitude of a Vector in $\mathbb{R}^3$?

Provide less scaffolding by asking students to determine an expression for the magnitude of $\boldsymbol{v}$ shown in the diagram below by simply showing the diagram on the board or document camera without the additional questions. Notice that $\mathbf{v}$ is a vector, but that x, y, z, and w are real numbers that represent lengths. For students who need a more concrete approach, label the terminal point of the vector $\mathbf{v}$ with an actual set of coordinates such as $(2,8,6)$ and then model how to find the magnitude of the vector before working through the problem in general.

- Use the Pythagorean theorem to write an equation relating x, y, and w.
  - $x^2+y^2=w^2$
- Use the Pythagorean theorem to write an equation relating w, z, and $\|\mathbf{v}\|$.
  - $w^2+z^2=\|\mathbf{v}\|^2$
- Now write an equation relating $\|\mathbf{v}\|$, x, y, and z.
  - $x^2+y^2=w^2$
  - $w^2=\|\mathbf{v}\|^2-z^2$
  - $x^2+y^2=\|\mathbf{v}\|^2-z^2$
  - $x^2+y^2+z^2=\|\mathbf{v}\|^2$

Example 2: What is the Magnitude of a Vector in $\mathbb{R}^3$?



MP.7

a. **Find a general formula for $\|\mathbf{v}\|^2$.**

$\|\mathbf{v}\|^2 = x^2 + y^2 + z^2$

b. **Solve this equation for $\|\mathbf{v}\|$ to find the magnitude of the vector.**

$\|\mathbf{v}\| = \sqrt{x^2 + y^2 + z^2}$

Exercises 5–8 (5 minutes)

In Exercise 5, students practice finding the magnitude of a vector in three dimensions. Direct students to work independently on Exercise 5 first, and then have them work with a partner to check their results.

Exercises 5–8

5. **Which vector has greater magnitude, $\mathbf{v} = \langle 0, 5, -4 \rangle$ or $\mathbf{u} = \langle 3, -4, 4 \rangle$? Show work to support your answer.**

$$\|\mathbf{v}\| = \sqrt{0^2 + 5^2 + (-4)^2} = \sqrt{41}$$

$$\|\mathbf{u}\| = \sqrt{3^2 + (-4)^2 + 4^2} = \sqrt{41}$$

These vectors have equal magnitude.

6. **Explain why vectors can have equal magnitude but not be the same vector.**

A vector has both a magnitude and a direction. If you graphed the vectors from Exercise 5, you can see they point in different directions so they cannot be the same even though they have equal magnitude.

Give students time to consider how the rules and representations learned in the previous lesson extend to three dimensions. If needed, scaffold this exercise by providing specific examples of vectors for students to work with (e.g., those shown in the table below) before asking them to write general rules.

	Vectors in $\mathbb{R}^2$	Vectors in $\mathbb{R}^3$
Component Form	$\langle 2,3 \rangle$	$\langle 2,3,4 \rangle$
Column Form	$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$
Magnitude	$\|\mathbf{v}\| = \sqrt{2^2 + 3^2} = \sqrt{13}$	$\|\mathbf{v}\| = \sqrt{2^2 + 3^2 + 4^2} = \sqrt{29}$
Addition	If $\mathbf{v} = \langle 2,3 \rangle$ and $\mathbf{u} = \langle 2, -4 \rangle$, Then $\mathbf{v} + \mathbf{u} = \langle 2 + 2, 3 + (-4) \rangle = \langle 4, -1 \rangle$	If $\mathbf{v} = \langle 2,3,4 \rangle$ and $\mathbf{u} = \langle 2, -4, 1 \rangle$, Then $\mathbf{v} + \mathbf{u} = \langle 2 + 2, 3 + (-4), 4 + 1 \rangle = \langle 4, -1, 5 \rangle$
Subtraction	If $\mathbf{v} = \langle 2,3 \rangle$ and $\mathbf{u} = \langle 2, -4 \rangle$, Then $\mathbf{v} + \mathbf{u} = \langle 2 - 2, 3 - (-4) \rangle = \langle 0,7 \rangle$	If $\mathbf{v} = \langle 2,3,4 \rangle$ and $\mathbf{u} = \langle 2, -4, 1 \rangle$, Then $\mathbf{v} + \mathbf{u} = \langle 2 - 2, 3 - (-4), 4 - 1 \rangle = \langle 0,7,3 \rangle$
Scalar Multiplication	If $\mathbf{v} = \langle 2,3 \rangle$ then $2\mathbf{v} = \langle 2 \cdot 2, 2 \cdot 3 \rangle = \langle 4,6 \rangle$	If $\mathbf{v} = \langle 2,3,4 \rangle$ then $2\mathbf{v} = \langle 2 \cdot 2, 2 \cdot 3, 2 \cdot 4 \rangle = \langle 4,6,8 \rangle$

7. Vector arithmetic in $\mathbb{R}^3$ is analogous to vector arithmetic in $\mathbb{R}^2$. Complete the graphic organizer to illustrate these ideas.

	Vectors in $\mathbb{R}^2$	Vectors in $\mathbb{R}^3$
Component Form	$\langle a, b \rangle$	$\langle a, b, c \rangle$
Column Form	$\begin{bmatrix} a \\ b \end{bmatrix}$	$\begin{bmatrix} a \\ b \\ c \end{bmatrix}$
Magnitude	$\|\mathbf{v}\| = \sqrt{a^2 + b^2}$	$\|\mathbf{v}\| = \sqrt{a^2 + b^2 + c^2}$
Addition	If $\mathbf{v} = \langle a, b \rangle$ and $\mathbf{u} = \langle c, d \rangle$, Then $\mathbf{v} + \mathbf{u} = \langle a + c, b + d \rangle$	*If* $\mathbf{v} = \langle a, b, c \rangle$ *and* $\mathbf{u} = \langle d, e, f \rangle$, *Then* $\mathbf{v} + \mathbf{u} = \langle a + d, b + e, c + f \rangle$
Subtraction	If $\mathbf{v} = \langle a, b \rangle$ and $\mathbf{u} = \langle c, d \rangle$, Then $\mathbf{v} - \mathbf{u} = \langle a - c, b - d \rangle$	*If* $\mathbf{v} = \langle a, b, c \rangle$ *and* $\mathbf{u} = \langle d, e, f \rangle$, *Then* $\mathbf{v} - \mathbf{u} = \langle a - d, b - e, c - f \rangle$
Scalar Multiplication	If $\mathbf{v} = \langle a, b \rangle$ and k is a real number $k\mathbf{v} = \langle ka, kb \rangle$	*If* $\mathbf{v} = \langle a, b, c \rangle$ *and* k *is a real number* $k\mathbf{v} = \langle ka, kb, kc \rangle$

Before starting Exercise 8, make sure students have correct information in the graphic organizer. Show a completed one to the class and, if time permits, have different groups explain how they got their answers for each row. Allow students time to correct any errors they may have made.

8. Given $\mathbf{v} = \langle 2, 0, -4 \rangle$ and $\mathbf{u} = \langle -1, 5, 3 \rangle$.

a. Calculate the following.

i. $\mathbf{v} + \mathbf{u}$

$\mathbf{v} + \mathbf{u} = \langle 2 + (-1), 0 + 5, -4 + 3 \rangle = \langle 1, 5, -1 \rangle$

ii. $2\mathbf{v} - \mathbf{u}$

$2\mathbf{v} - \mathbf{u} = \langle 2 \cdot 2 - (-1), 2 \cdot 0 - 5, 2 \cdot (-4) - 3 \rangle = \langle 5, -5, -11 \rangle$

iii. $\|\mathbf{u}\|$

$\|\mathbf{u}\| = \sqrt{(-1)^2 + 5^2 + 3^2} = \sqrt{35}$

b. Suppose the point $(1, 3, 5)$ is translated by $\mathbf{v}$ and then by $\mathbf{u}$. Determine a vector $\mathbf{w}$ that would return the point back to its original location $(1, 3, 5)$.

From part (a), we have $\mathbf{v} + \mathbf{u} = \langle 1, 5, -1 \rangle$. The point will be translated from $(1, 3, 5)$ to $(2, 8, 4)$ since $1 + 1 = 2$, $5 + 3 = 8$, and $5 - 1 = 4$. The vector that will return this point to its original location will be the opposite of $\mathbf{v} + \mathbf{u}$.

$$-(\mathbf{v} + \mathbf{u}) = \langle -1, -5, 1 \rangle$$

And this vector will translate $(2, 8, 4)$ back to $(1, 3, 5)$ because $2 - 1 = 1$, $8 - 5 = 3$, and $4 + 1 = 5$.

Closing (2 minutes)

Give students time to respond to the questions below individually in writing or with a partner.

- Why can we represent vectors as translation maps?
  - *They mean the same thing geometrically.*
- How are two-dimensional vectors the same as three-dimensional vectors, and how are they different?
  - *Either dimension has a magnitude and direction and represents a translation or shift. They are different because we need one additional component to describe a vector in three dimensions.*

The graphic organizer in Exercise 7 can serve as a lesson summary of vector arithmetic in three dimensions along with the information about translation maps shown below.

Lesson Summary

A vector $\mathbf{v}$ can define a translation map $T_{\mathbf{v}}$ that takes a point to its image under the translation. Applying the map to the set of all points that make up a geometric figure serves to translate the figure by the vector.

Exit Ticket (3 minutes)

Name _____ Date _____

Vectors and Translation Maps

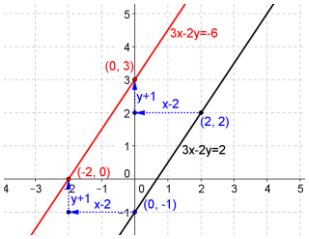
Exit Ticket

1. Given the vector $\mathbf{v} = \langle -2,1 \rangle$, find the image of the line $3x - 2y = 2$ under the translation map $T_{\mathbf{v}}$. Graph the original line and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$.

2. Given the vector $\mathbf{v} = \langle -1,2 \rangle$, find the image of the circle $(x - 2)^2 + (y + 1)^2 = 4$ under the translation map $T_{\mathbf{v}}$. Graph the original circle and its image, and then explain the geometric effect of the map $T_{\mathbf{v}}$.

Exit Ticket Sample Solutions

1. Given the vector $\mathbf{v} = \langle -2, 1 \rangle$, find the image of the line $3x - 2y = 2$ under the translation map $T_{\mathbf{v}}$. Graph the original line and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$.

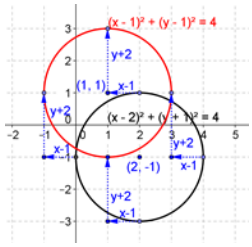


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{pmatrix} x-2 \\ y+1 \end{pmatrix}$; *image of the original line:* $3(x+2) - 2(y-1) = 2$.

$3x + 6 - 2y + 2 = 2,\ 3x - 2y = -6$

Every point on the line is shifted 2 units to the left and 1 unit upward.

2. Given the vector $\mathbf{v} = \langle -1, 2 \rangle$, find the image of the circle $(x-2)^2 + (y+1)^2 = 4$ under the translation map $T_{\mathbf{v}}$. Graph the original circle and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$.



$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{pmatrix} x-1 \\ y+2 \end{pmatrix}$; *image of original circle:* $(x-1)^2 + (y-1)^2 = 4$.

Every point on the circle is shifted one unit to the left and two units upward. The new center is $(1, 1)$ *and the radius* $r = 2$ *stays the same.*

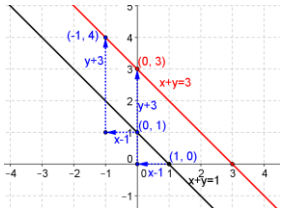
Problem Set Sample Solutions

1. Myishia says that when applying the translation map $T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y-2 \end{bmatrix}$ to a set of points given by an equation relating x and y, we should replace every x that is in the equation by $x + 1$, and y by $y - 2$. For example, the equation of the parabola $y = x^2$ would become $y - 2 = (x+1)^2$. Is she correct? Explain your answer.

No, she is not correct. What she did translates the points on the parabola $y = x^2$ *in the opposite direction—one unit to the left and 2 units upward, which is not the geometric effect of* $T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y-2 \end{bmatrix}$. *In order to have the correct translation based on* $T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y-2 \end{bmatrix}$, *we need to set* $x' = x + 1$ *and* $y' = y - 2$, *which is equivalent to* $x = x' - 1$ *and* $y = y' + 2$.

This process gives the transformed equation $y' + 2 = (x' - 1)^2$, *which we write as* $y + 2 = (x-1)^2$.

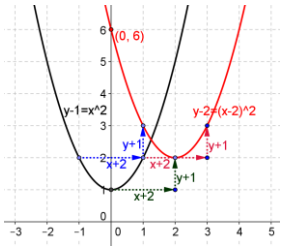
2. Given the vector $\mathbf{v} = \langle -1, 3 \rangle$, find the image of the line $x + y = 1$ under the translation map $T_{\mathbf{v}}$. Graph the original line and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the line.



$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x-1 \\ y+3 \end{bmatrix}$, $(x-(-1)) + (y-(3)) = 1$, $x+1+y-3=1$, $x+y=3$

Every point on the line is shifted one unit left and three units upward. The slopes of the lines remain -1.

3. Given the vector $\mathbf{v} = \langle 2, 1 \rangle$, find the image of the parabola $y - 1 = x^2$ under the translation map $T_{\mathbf{v}}$. Draw a graph of the original parabola and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the parabola. Find the vertex and x-intercepts of the graph of the image.

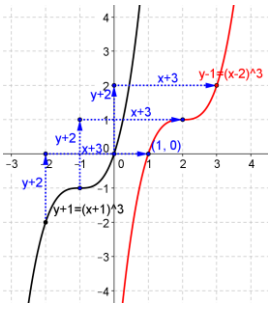


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+2 \\ y+1 \end{bmatrix}$, $y - 2 = (x-2)^2$

Every point on the parabola is shifted two units to the right and one unit upward.

The vertex is $(2, 2)$, and there are no x-intercepts.

4. Given the vector $\mathbf{v} = \langle 3, 2 \rangle$, find the image of the graph of $y + 1 = (x+1)^3$ under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the x-intercepts of the graph of the image.

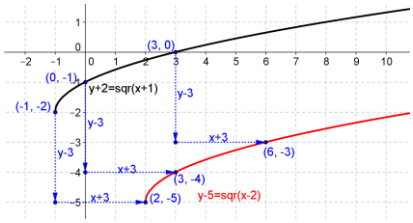


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+3 \\ y+2 \end{bmatrix}$, $y - 1 = (x-2)^3$

Every point on the curve is shifted three units to the right and two units upward.

The x-intercept is 1.

5. Given the vector $\mathbf{v} = \langle 3, -3 \rangle$, find the image of the graph of $y + 2 = \sqrt{x+1}$ under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the x-intercepts of the graph of the image.

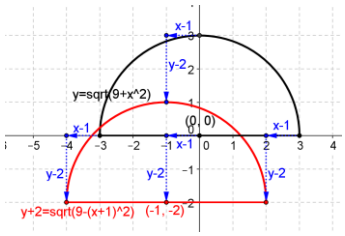


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+3 \\ y-3 \end{bmatrix}$, $y + 5 = \sqrt{x-2}$

Every point on the curve is shifted three units to the right and three units downward.

The x-intercept is 27.

6. Given the vector $\mathbf{v} = \langle -1, -2 \rangle$, find the image of the graph of $y = \sqrt{9 - x^2}$ under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the x-intercepts of the graph of the image.

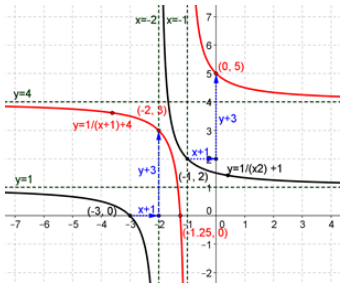


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x-1 \\ y-2 \end{bmatrix}$, $y + 2 = \sqrt{9 - (x+1)^2}$

Every point on the semicircle is shifted one unit to the left and two units downward.

x-intercepts: $-1 + \sqrt{5}$ and $-1 - \sqrt{5}$.

7. Given the vector $\mathbf{v} = \langle 1, 3 \rangle$, find the image of the graph of $y = \dfrac{1}{x+2} + 1$ under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the equations of the asymptotes of the graph of the image.

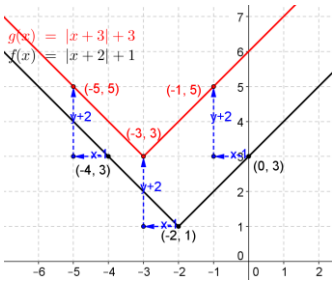


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y+3 \end{bmatrix}$, $y = \dfrac{1}{x+1} + 4$

Every point on the curve is shifted one unit to the right and three units upward.

The new vertical asymptote is $x = -1$, the new horizontal asymptote is $y = 4$.

8. **Given the vector $\mathbf{v} = \langle -1, 2 \rangle$, find the image of the graph of $y = |x + 2| + 1$ under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the x-intercepts of the graph of the image.**

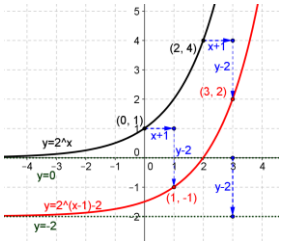


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - 1 \\ y + 2 \end{bmatrix}$, $y = |x + 3| + 3$

Every point on the graph is shifted one unit to the left and two units upward.

There are no x-intercepts.

9. **Given the vector $\mathbf{v} = \langle 1, -2 \rangle$, find the image of the graph of $y = 2^x$ under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the x-intercepts of the graph of the image.**

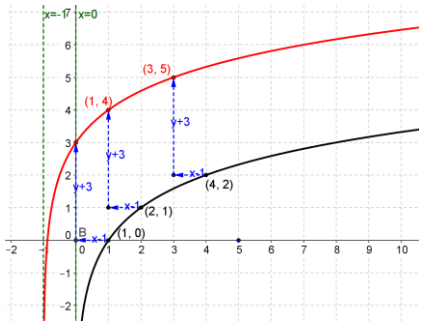


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + 1 \\ y - 2 \end{bmatrix}$, $y = 2^{x-1} - 2$

Every point on the curve is shifted one unit to the right and two units downward.

x-intercept is 2.

10. Given the vector $\mathbf{v} = \langle -1, 3 \rangle$, find the image of the graph of $y = \log_2 x$, under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the x-intercepts of the graph of the image.

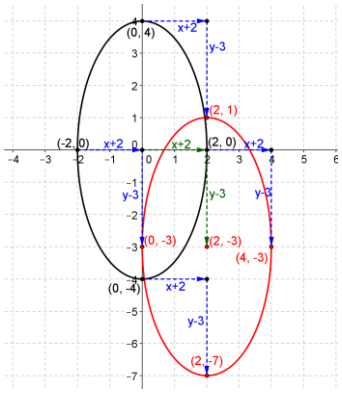


$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x-1 \\ y+3 \end{bmatrix}$, $y = \log_2(x+1) + 3$

Every point on the curve is shifted one unit to the left and three units upward.

x-intercept: $-\frac{7}{8}$.

11. Given the vector $v = \langle 2, -3 \rangle$, find the image of the graph of $\frac{x^2}{4} + \frac{y^2}{16} = 1$ under the translation map $T_{\mathbf{v}}$. Draw the original graph and its image, and explain the geometric effect of the map $T_{\mathbf{v}}$ on the graph. Find the new center, major and minor axis of the graph of the image.



$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+2 \\ y-3 \end{bmatrix}$, $\frac{(x-2)^2}{4} + \frac{(y+3)^2}{16} = 1$

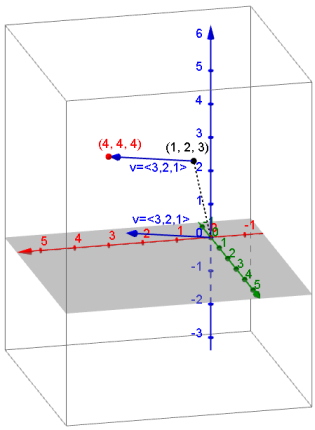
Every point on the ellipse is shifted two units to the right and three units downward.

The new center is $(2, -3)$*, the major axis is* 4*, and the minor axis is* 2.

12. Given the vector v, find the image of the given point P under the translation map $T_{\mathbf{v}}$. Graph P and its image.

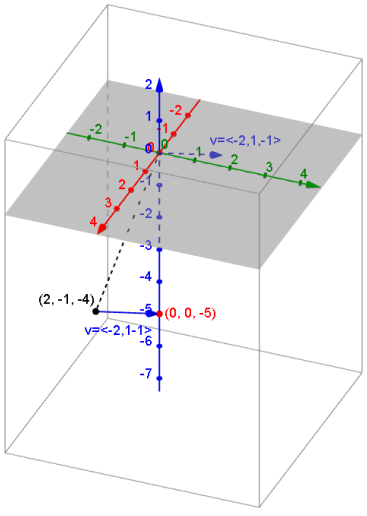
a. $\mathbf{v} = \langle 3, 2, 1 \rangle$, $P = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+3 \\ y+2 \\ z+1 \end{bmatrix}$, *the new image:* $\begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$



b. $\mathbf{v} = \langle -2, 1, -1 \rangle$, $P = \begin{bmatrix} 2 \\ -1 \\ -4 \end{bmatrix}$

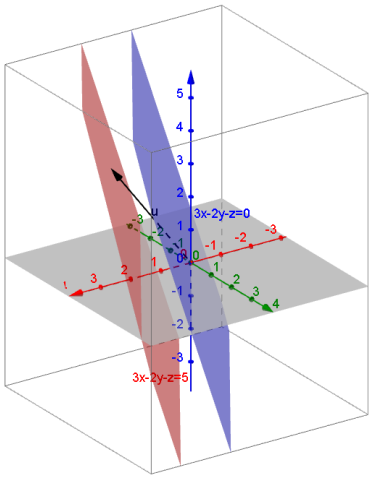
$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-2 \\ y+1 \\ z-1 \end{bmatrix}$, *the new image:* $\begin{bmatrix} 0 \\ 0 \\ -5 \end{bmatrix}$



13. Given the vector v, find the image of the given plane under the translation map $T_{\mathbf{v}}$. Sketch the original vector and its image.

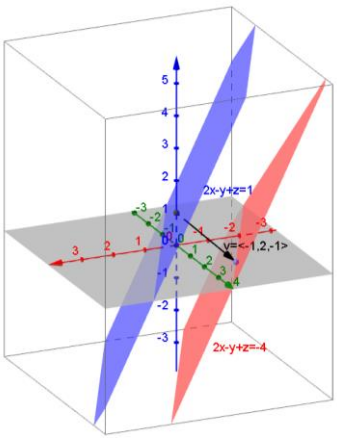
a. $\mathbf{v} = \langle 2, -1, 3 \rangle,\ 3x - 2y - z = 0$

$$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+2 \\ y-1 \\ z+3 \end{bmatrix},\ 3x - 2y - z = 5$$



b. $\mathbf{v} = \langle -1, 2, -1 \rangle,\ 2x - y + z = 1$

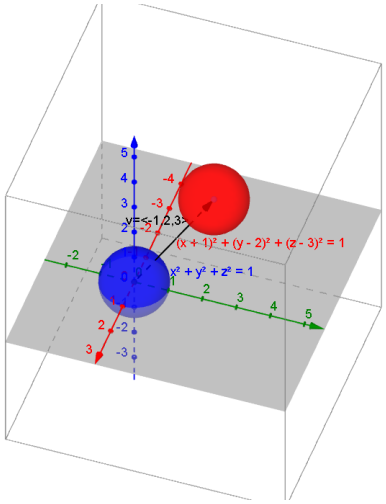
$$T_{\mathbf{v}}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-1 \\ y+2 \\ z-1 \end{bmatrix}, 2x - y + z = -4$$



14. Given the vector v, find the image of the given sphere under the translation map T_v. Sketch the original sphere and its image.

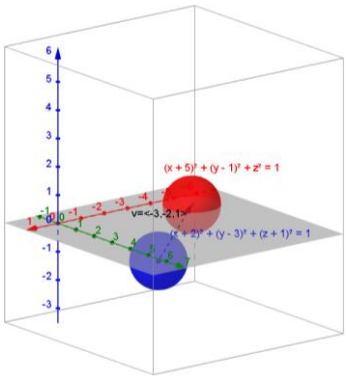
a. $\mathbf{v} = \langle -1, 2, 3 \rangle,\ x^2 + y^2 + z^2 = 1$

$$T_v\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-1 \\ y+2 \\ z+3 \end{bmatrix}, (x+1)^2 + (y-2)^2 + (z-3)^2 = 1$$



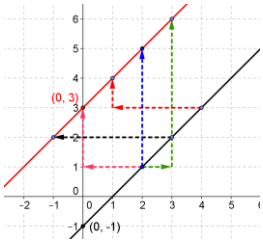
b. $\mathbf{v} = \langle -3, -2, 1 \rangle,\ (x+2)^2 + (y-3)^2 + (z+1)^2 = 1$

$$T_v\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-3 \\ y-2 \\ z+1 \end{bmatrix}, (x+5)^2 + (y-1)^2 + (z)^2 = 1$$



15. Find a vector v and translation map T_v that will translate the line $x - y = 1$ to the line $x - y = -3$. Sketch the original vector and its image.

Answers vary. For example, $\mathbf{v} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$ *and* $T_v\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+0 \\ y+4 \end{bmatrix}$.



16. Find a vector v and translation map T_v that will translate the parabola $y = x^2 + 4x + 1$ to the parabola $y = x^2$

Because $y = x^2 + 4x + 1$ *can be written as* $y + 3 = (x+2)^2$, $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ *and* $T_v\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+2 \\ y+3 \end{bmatrix}$.

17. Find a vector v and translation map T_v that will translate the circle with equation $x^2 + y^2 - 4x + 2y - 4 = 0$ to the circle with equation $(x+3)^2 + (y-4)^2 = 9$

Because $x^2 - 4x + y^2 + 2y = 4$ *can be written as* $(x-2)^2 + (y+1)^2 = 9$, $\mathbf{v} = \begin{bmatrix} -5 \\ 5 \end{bmatrix}$ *and* $T_v\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x-5 \\ y+5 \end{bmatrix}$.

18. Find a vector v and translation map T_v that will translate the graph of $y = \sqrt{x-3} + 2$ to the graph of $y = \sqrt{x+2} - 3$.

$\mathbf{v} = \begin{bmatrix} -5 \\ -5 \end{bmatrix}$ *and* $T_v\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x-5 \\ y-5 \end{bmatrix}$

19. Find a vector v and translation map T_v that will translate the sphere $(x+2)^2 + (y-3)^2 + (z+1)^2 = 1$ to the sphere $(x-3)^2 + (y+1)^2 + (z+2)^2 = 1$

$\mathbf{v} = \begin{bmatrix} 5 \\ -4 \\ -1 \end{bmatrix}$ *and* $T_v\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+5 \\ y-4 \\ z-1 \end{bmatrix}$

20. Given vectors $\mathbf{u} = \langle 2, -1, 3 \rangle$, $\mathbf{v} = \langle 2, 0, -2 \rangle$, and $\mathbf{w} = \langle -3, 6, 0 \rangle$, find the following.

a. $3\mathbf{u} + \mathbf{v} + \mathbf{w}$

$\langle 5, 3, 7 \rangle$

b. $\mathbf{w} - 2\mathbf{v} - \mathbf{u}$

$\langle -9, 7, 1 \rangle$

c. $3\left(2\mathbf{u} - \frac{1}{2}\mathbf{v}\right) - \frac{1}{3}\mathbf{w}$

$\langle 10, -8, 21 \rangle$

d. $-2\mathbf{u} - 3(5\mathbf{v} - 3\mathbf{w})$

$\langle -61, 56, 24 \rangle$

e. $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{w}\|$

$\|\mathbf{u}\| = \sqrt{14}, \|\mathbf{v}\| = 2\sqrt{2}, \|\mathbf{w}\| = \sqrt{45}$

f. Show that $2\|\mathbf{v}\| = \|2\mathbf{v}\|$.

$2\|\mathbf{v}\| = 2\left(2\sqrt{2}\right) = 4\sqrt{2}, \ \|2\mathbf{v}\| = \|\langle 4, 0, -4\rangle\| = \sqrt{32} = 4\sqrt{2}$

g. Show that $\|\mathbf{u} + \mathbf{v}\| \neq \|\mathbf{u}\| + \|\mathbf{v}\|$.

$\|\mathbf{u} + \mathbf{v}\| = \|\langle 4, -1, 1\rangle\| = \sqrt{18}, \ \|\mathbf{u}\| + \|\mathbf{v}\| = \sqrt{14} + 2\sqrt{2}$

h. Show that $\|\mathbf{v} - \mathbf{w}\| \neq \|\mathbf{v}\| - \|\mathbf{w}\|$.

$\|\mathbf{v} - \mathbf{w}\| = \|\langle 5, -6, -2\rangle\| = \sqrt{65}, \ \|\mathbf{v}\| - \|\mathbf{w}\| = 2\sqrt{2} - \sqrt{45}$

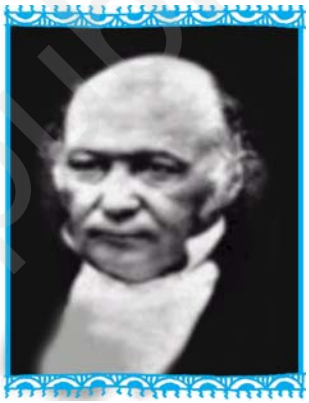
i. $\frac{1}{\|\mathbf{u}\|}\mathbf{u}$ and $\left\|\frac{1}{\|\mathbf{u}\|}\mathbf{u}\right\|$

$$\frac{1}{\|\mathbf{u}\|}\mathbf{u} = \frac{\langle 2, -1, 3\rangle}{\sqrt{14}} = \left\langle \frac{2}{\sqrt{14}}, -\frac{1}{\sqrt{14}}, \frac{3}{\sqrt{14}} \right\rangle, \quad \left\|\frac{1}{\|\mathbf{u}\|}\mathbf{u}\right\| = \sqrt{\left(\frac{2}{\sqrt{14}}\right)^2 + \left(-\frac{1}{\sqrt{14}}\right)^2 + \left(\frac{3}{\sqrt{14}}\right)^2} = \sqrt{\frac{4}{14} + \frac{1}{14} + \frac{9}{14}}$$
$$= 1$$

Chapter Five: Vector Algebra

❖ *In most sciences one generation tears down what another has built and what one has established another undoes. In Mathematics alone each generation builds a new story to the old structure. – HERMAN HANKEL* ❖

Introduction



W.R. Hamilton (1805-1865)

In our day to day life, we come across many queries such as – What is your height? How should a football player hit the ball to give a pass to another player of his team? Observe that a possible answer to the first query may be 1.6 meters, a quantity that involves only one value (magnitude) which is a real number. Such quantities are called *scalars*. However, an answer to the second query is a quantity (called force) which involves muscular strength (magnitude) and direction (in which another player is positioned). Such quantities are called *vectors.* In mathematics, physics and engineering, we frequently come across with both types of quantities, namely, scalar quantities such as length, mass, time, distance, speed, area, volume, temperature, work, money, voltage, density, resistance etc. and vector quantities like displacement, velocity, acceleration, force, weight, momentum, electric field intensity etc.

In this chapter, we will study some of the basic concepts about vectors, various operations on vectors, and their algebraic and geometric properties. These two type of properties, when considered together give a full realisation to the concept of vectors, and lead to their vital applicability in various areas as mentioned above.

10.2 Some Basic Concepts

Let '*l*' be any straight line in plane or three dimensional space. This line can be given two directions by means of arrowheads. A line with one of these directions prescribed is called a *directed line* (Fig 10.1 (i), (ii)).

© NCERT not to be republished

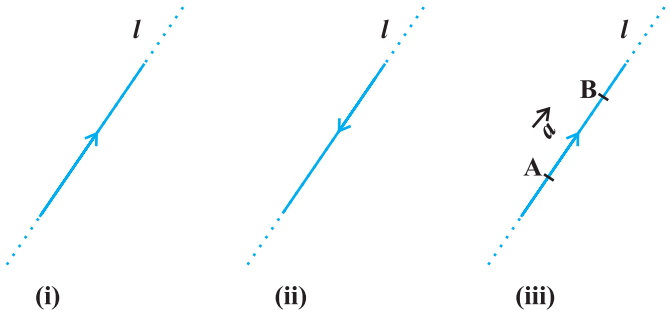


Fig 10.1

Now observe that if we restrict the line *l* to the line segment AB, then a magnitude is prescribed on the line *l* with one of the two directions, so that we obtain a *directed line segment* (Fig 10.1(iii)). Thus, a directed line segment has magnitude as well as direction.

Definition 1 A quantity that has magnitude as well as direction is called a vector.

Notice that a directed line segment is a vector (Fig 10.1(iii)), denoted as $\overrightarrow{AB}$ or simply as $\vec{a}$, and read as 'vector $\overrightarrow{AB}$' or 'vector $\vec{a}$'.

The point A from where the vector $\overrightarrow{AB}$ starts is called its *initial point*, and the point B where it ends is called its *terminal point*. The distance between initial and terminal points of a vector is called the *magnitude* (or length) of the vector, denoted as $|\overrightarrow{AB}|$, or $|\vec{a}|$, or *a*. The arrow indicates the direction of the vector.

Note Since the length is never negative, the notation $|\vec{a}| < 0$ has no meaning.

Position Vector

From Class XI, recall the three dimensional right handed rectangular coordinate system (Fig 10.2(i)). Consider a point P in space, having coordinates (*x*, *y*, *z*) with respect to the origin O(0, 0, 0). Then, the vector $\overrightarrow{OP}$ having O and P as its initial and terminal points, respectively, is called the *position vector* of the point P with respect to O. Using distance formula (from Class XI), the magnitude of $\overrightarrow{OP}$ (or $\vec{r}$) is given by

$$|\overrightarrow{OP}| = \sqrt{x^2 + y^2 + z^2}$$

In practice, the position vectors of points A, B, C, etc., with respect to the origin O are denoted by $\vec{a}$, $\vec{b}, \vec{c}$, etc., respectively (Fig 10.2 (ii)).

© NCERT not to be republished

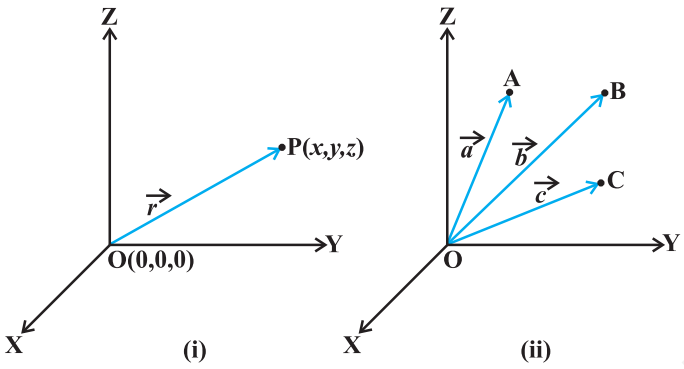


Fig 10.2

Direction Cosines

Consider the position vector $\overrightarrow{OP}$ or $\vec{r}$ of a point P(x, y, z) as in Fig 10.3. The angles α, β, γ made by the vector $\vec{r}$ with the positive directions of x, y and z-axes respectively, are called its *direction angles*. The cosine values of these angles, i.e., $\cos\alpha$, $\cos\beta$ and $\cos\gamma$ are called *direction cosines* of the vector $\vec{r}$, and usually denoted by l, m and n, respectively.

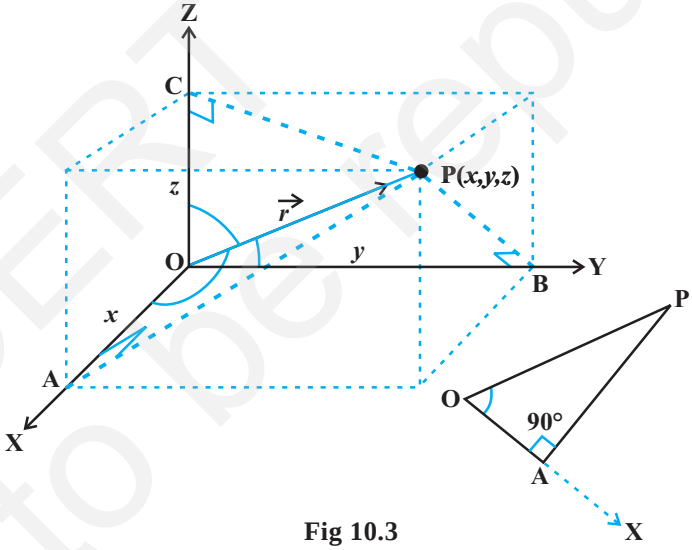


Fig 10.3

From Fig 10.3, one may note that the triangle OAP is right angled, and in it, we have $\cos\alpha = \frac{x}{r}$ (r stands for $|\vec{r}|$). Similarly, from the right angled triangles OBP and OCP, we may write $\cos\beta = \frac{y}{r}$ and $\cos\gamma = \frac{z}{r}$. Thus, the coordinates of the point P may also be expressed as (lr, mr,nr). The numbers lr, mr and nr, proportional to the direction cosines are called as *direction ratios* of vector $\vec{r}$, and denoted as a, b and c, respectively.

© NCERT not to be republished

Note One may note that $l^2 + m^2 + n^2 = 1$ but $a^2 + b^2 + c^2 \neq 1$, in general.

Types of Vectors

Zero Vector A vector whose initial and terminal points coincide, is called a zero vector (or null vector), and denoted as $\vec{0}$. Zero vector can not be assigned a definite direction as it has zero magnitude. Or, alternatively otherwise, it may be regarded as having any direction. The vectors $\overrightarrow{AA}, \overrightarrow{BB}$ represent the zero vector,

Unit Vector A vector whose magnitude is unity (i.e., 1 unit) is called a unit vector. The unit vector in the direction of a given vector $\vec{a}$ is denoted by $\hat{a}$.

Coinitial Vectors Two or more vectors having the same initial point are called coinitial vectors.

Collinear Vectors Two or more vectors are said to be collinear if they are parallel to the same line, irrespective of their magnitudes and directions.

Equal Vectors Two vectors $\vec{a}$ and $\vec{b}$ are said to be equal, if they have the same magnitude and direction regardless of the positions of their initial points, and written as $\vec{a} = \vec{b}$.

Negative of a Vector A vector whose magnitude is the same as that of a given vector (say, $\overrightarrow{AB}$), but direction is opposite to that of it, is called *negative* of the given vector. For example, vector $\overrightarrow{BA}$ is negative of the vector $\overrightarrow{AB}$, and written as $\overrightarrow{BA} = -\overrightarrow{AB}$.

Remark The vectors defined above are such that any of them may be subject to its parallel displacement without changing its magnitude and direction. Such vectors are called *free vectors*. Throughout this chapter, we will be dealing with free vectors only.

Example 1 Represent graphically a displacement of 40 km, 30° west of south.

Solution The vector $\overrightarrow{OP}$ represents the required displacement (Fig 10.4).

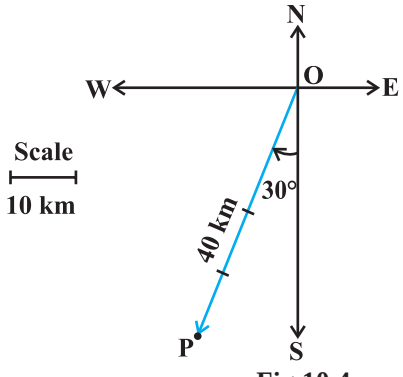


Fig 10.4

Example 2 Classify the following measures as scalars and vectors.

(i) 5 seconds

(ii) 1000 cm^3

(iii) 10 Newton (iv) 30 km/hr (v) 10 g/cm^3

(vi) 20 m/s towards north

Solution

(i) Time-scalar (ii) Volume-scalar (iii) Force-vector

(iv) Speed-scalar (v) Density-scalar (vi) Velocity-vector

Example 3 In Fig 10.5, which of the vectors are:

(i) Collinear (ii) Equal (iii) Coinitial

Solution

(i) Collinear vectors : $\vec{a}$, $\vec{c}$ and $\vec{d}$.

(ii) Equal vectors : $\vec{a}$ and $\vec{c}$.

(iii) Coinitial vectors : $\vec{b}$, $\vec{c}$ and $\vec{d}$.

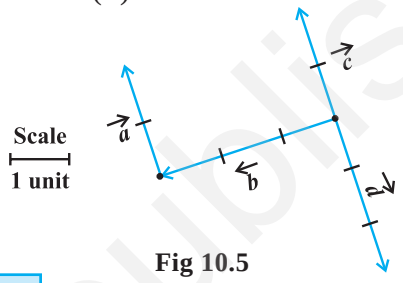


Fig 10.5

EXERCISE 10.1

1. Represent graphically a displacement of 40 km, 30° east of north.
2. Classify the following measures as scalars and vectors.

   (i) 10 kg (ii) 2 meters north-west (iii) 40°

   (iv) 40 watt (v) 10^{-19} coulomb (vi) 20 m/s^2
3. Classify the following as scalar and vector quantities.

   (i) time period (ii) distance (iii) force

   (iv) velocity (v) work done
4. In Fig 10.6 (a square), identify the following vectors.

   (i) Coinitial (ii) Equal

   (iii) Collinear but not equal

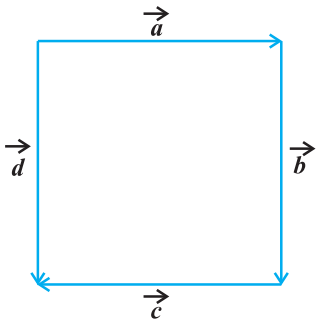


Fig 10.6

5. Answer the following as true or false.

   (i) $\vec{a}$ and $-\vec{a}$ are collinear.

   (ii) Two collinear vectors are always equal in magnitude.

   (iii) Two vectors having same magnitude are collinear.

   (iv) Two collinear vectors having the same magnitude are equal.

© NCERT not to be republished

Addition of Vectors

A vector $\overrightarrow{AB}$ simply means the displacement from a point A to the point B. Now consider a situation that a girl moves from A to B and then from B to C (Fig 10.7). The net displacement made by the girl from point A to the point C, is given by the vector $\overrightarrow{AC}$ and expressed as

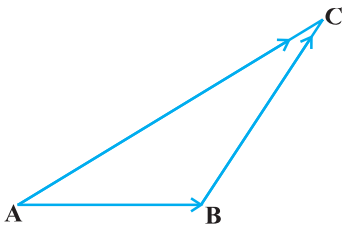


Fig 10.7

$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC}$$

This is known as the *triangle law of vector addition.*

In general, if we have two vectors $\vec{a}$ and $\vec{b}$ (Fig 10.8 (i)), then to add them, they are positioned so that the initial point of one coincides with the terminal point of the other (Fig 10.8(ii)).

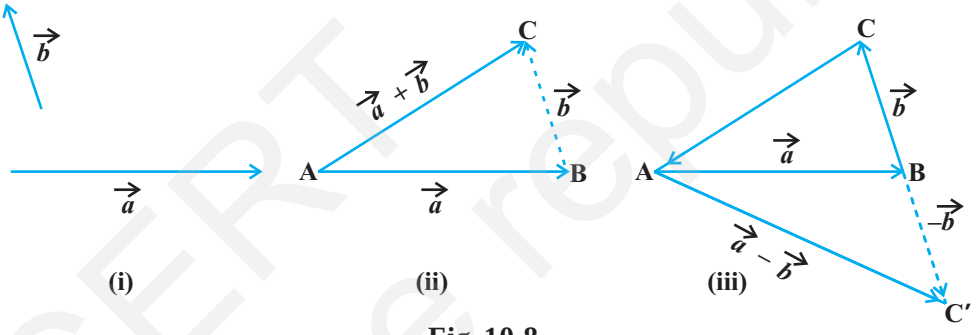


Fig 10.8

For example, in Fig 10.8 (ii), we have shifted vector $\vec{b}$ without changing its magnitude and direction, so that it's initial point coincides with the terminal point of $\vec{a}$. Then, the vector $\vec{a} + \vec{b}$, represented by the third side AC of the triangle ABC, gives us the sum (or resultant) of the vectors $\vec{a}$ and $\vec{b}$ i.e., in triangle ABC (Fig 10.8 (ii)), we have

$$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$$

Now again, since $\overrightarrow{AC} = -\overrightarrow{CA}$, from the above equation, we have

$$\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA} = \overrightarrow{AA} = \vec{0}$$

This means that when the sides of a triangle are taken in order, it leads to zero resultant as the initial and terminal points get coincided (Fig 10.8(iii)).

© NCERT not to be republished

Now, construct a vector $\overrightarrow{BC'}$ so that its magnitude is same as the vector $\overrightarrow{BC}$, but the direction opposite to that of it (Fig 10.8 (iii)), i.e.,

$$\overrightarrow{BC'} = -\overrightarrow{BC}$$

Then, on applying triangle law from the Fig 10.8 (iii), we have

$$\overrightarrow{AC'} = \overrightarrow{AB} + \overrightarrow{BC'} = \overrightarrow{AB} + (-\overrightarrow{BC}) = \vec{a} - \vec{b}$$

The vector $\overrightarrow{AC'}$ is said to represent the *difference of* $\vec{a}$ and $\vec{b}$.

Now, consider a boat in a river going from one bank of the river to the other in a direction perpendicular to the flow of the river. Then, it is acted upon by two velocity vectors–one is the velocity imparted to the boat by its engine and other one is the velocity of the flow of river water. Under the simultaneous influence of these two velocities, the boat in actual starts travelling with a different velocity. To have a precise idea about the effective speed and direction (i.e., the resultant velocity) of the boat, we have the following law of vector addition.

If we have two vectors $\vec{a}$ and $\vec{b}$ represented by the two adjacent sides of a parallelogram in magnitude and direction (Fig 10.9), then their sum $\vec{a} + \vec{b}$ is represented in magnitude and direction by the diagonal of the parallelogram through their common point. This is known as the *parallelogram law of vector addition.*

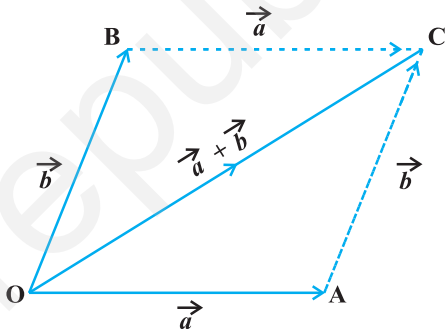


Fig 10.9

Note From Fig 10.9, using the triangle law, one may note that

$$\overrightarrow{OA} + \overrightarrow{AC} = \overrightarrow{OC}$$

or $$\overrightarrow{OA} + \overrightarrow{OB} = \overrightarrow{OC} \qquad (\text{since } \overrightarrow{AC} = \overrightarrow{OB})$$

which is parallelogram law. Thus, we may say that the two laws of vector addition are equivalent to each other.

Properties of vector addition

Property 1 For any two vectors $\vec{a}$ and $\vec{b}$,

$$\vec{a} + \vec{b} = \vec{b} + \vec{a} \qquad \text{(Commutative property)}$$

© NCERT not to be republished

Proof Consider the parallelogram ABCD (Fig 10.10). Let $\overrightarrow{AB}$ $\vec{a}$ and $\overrightarrow{BC}$ $\vec{b}$, then using the triangle law, from triangle ABC, we have

$$\overrightarrow{AC} = \vec{a} + \vec{b}$$

Now, since the opposite sides of a parallelogram are equal and parallel, from Fig 10.10, we have, $\overrightarrow{AD} = \overrightarrow{BC} = \vec{b}$ and $\overrightarrow{DC} = \overrightarrow{AB} = \vec{a}$. Again using triangle law, from triangle ADC, we have

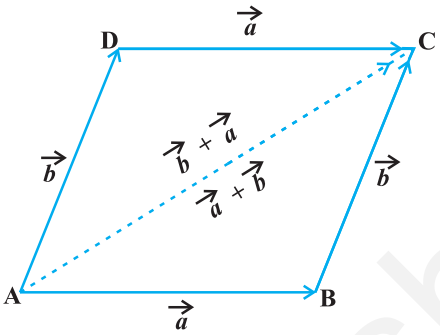


Fig 10.10

$$\overrightarrow{AC} = \overrightarrow{AD} + \overrightarrow{DC} = \vec{b} + \vec{a}$$

Hence $\vec{a} + \vec{b} = \vec{b} + \vec{a}$

Property 2 For any three vectors $\vec{a}, \vec{b}$ and $\vec{c}$

$$(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$$ (Associative property)

Proof Let the vectors $\vec{a}, \vec{b}$ and $\vec{c}$ be represented by $\overrightarrow{PQ}$, $\overrightarrow{QR}$ and $\overrightarrow{RS}$, respectively, as shown in Fig 10.11(i) and (ii).

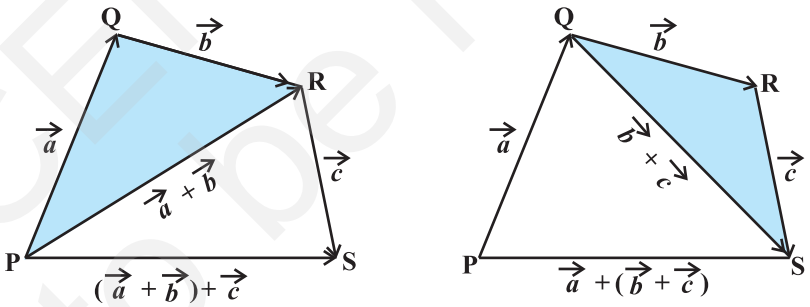


Fig 10.11

Then $\vec{a} + \vec{b} = \overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$

and $\vec{b} + \vec{c} = \overrightarrow{QR} + \overrightarrow{RS} = \overrightarrow{QS}$

So $(\vec{a} + \vec{b}) + \vec{c} = \overrightarrow{PR} + \overrightarrow{RS} = \overrightarrow{PS}$

© NCERT not to be republished

and $\vec{a} + (\vec{b} + \vec{c}) = \overrightarrow{PQ} + \overrightarrow{QS} = \overrightarrow{PS}$

Hence $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$

Remark The associative property of vector addition enables us to write the sum of three vectors $\vec{a}$, $\vec{b}$, $\vec{c}$ as $\vec{a} + \vec{b} + \vec{c}$ without using brackets.

Note that for any vector $\vec{a}$, we have

$$\vec{a} + \vec{0} = \vec{0} + \vec{a} = \vec{a}$$

Here, the zero vector $\vec{0}$ is called the *additive identity* for the vector addition.

10.5 Multiplication of a Vector by a Scalar

Let $\vec{a}$ be a given vector and λ a scalar. Then the product of the vector $\vec{a}$ by the scalar λ, denoted as $\lambda\vec{a}$, is called the multiplication of vector $\vec{a}$ by the scalar λ. Note that, $\lambda\vec{a}$ is also a vector, collinear to the vector $\vec{a}$. The vector $\lambda\vec{a}$ has the direction same (or opposite) to that of vector $\vec{a}$ according as the value of λ is positive (or negative). Also, the magnitude of vector $\lambda\vec{a}$ is $|\lambda|$ times the magnitude of the vector $\vec{a}$, i.e.,

$$|\lambda\vec{a}| = |\lambda||\vec{a}|$$

A geometric visualisation of multiplication of a vector by a scalar is given in Fig 10.12.

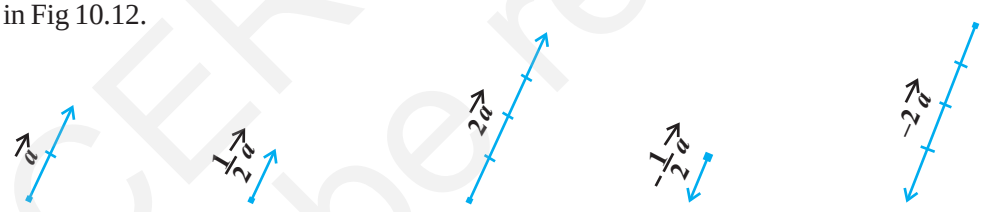


Fig 10.12

When $\lambda = -1$, then $\lambda\vec{a} = -\vec{a}$, which is a vector having magnitude equal to the magnitude of $\vec{a}$ and direction opposite to that of the direction of $\vec{a}$. The vector $-\vec{a}$ is called the *negative* (or *additive inverse*) *of vector* $\vec{a}$ and we always have

$$\vec{a} + (-\vec{a}) = (-\vec{a}) + \vec{a} = \vec{0}$$

Also, if $\lambda = \frac{1}{|\vec{a}|}$, provided $\vec{a}$ 0, i.e. $\vec{a}$ is not a null vector, then

$$|\lambda\vec{a}| = |\lambda||\vec{a}| = \frac{1}{|\vec{a}|}|\vec{a}| \quad 1$$

© NCERT not to be republished

So, $\lambda \vec{a}$ *represents the unit vector in the direction of* $\vec{a}$. We write it as

$$\hat{a} = \frac{1}{|\vec{a}|}\vec{a}$$

Note For any scalar k, $k\vec{0} = \vec{0}$.

10.5.1 *Components of a vector*

Let us take the points A(1, 0, 0), B(0, 1, 0) and C(0, 0, 1) on the *x*-axis, *y*-axis and *z*-axis, respectively. Then, clearly

$$|\overrightarrow{OA}| = 1, |\overrightarrow{OB}| = 1 \text{ and } |\overrightarrow{OC}| = 1$$

The vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$, each having magnitude 1, are called *unit vectors along the axes* OX, OY and OZ, respectively, and denoted by $\hat{i}$, $\hat{j}$ and $\hat{k}$, respectively (Fig 10.13).

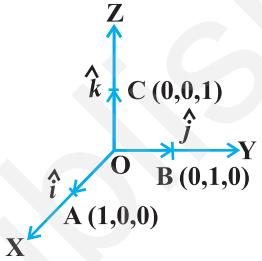


Fig 10.13

Now, consider the position vector $\overrightarrow{OP}$ of a point P(*x*, *y*, *z*) as in Fig 10.14. Let P_1 be the foot of the perpendicular from P on the plane XOY. We, thus, see that P_1P is

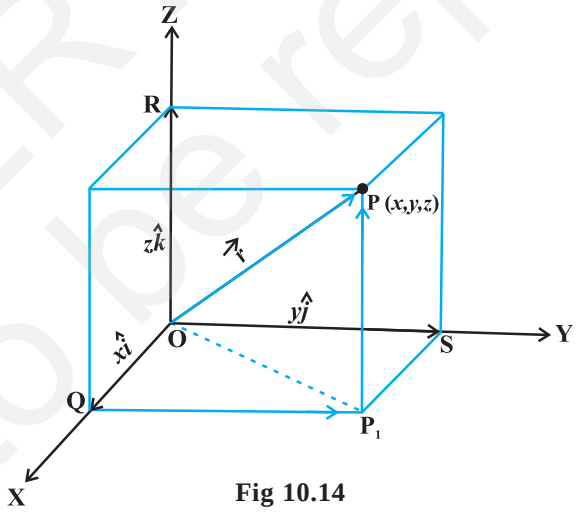


Fig 10.14

parallel to *z*-axis. As $\hat{i}$, $\hat{j}$ and $\hat{k}$ are the unit vectors along the *x*, *y* and *z*-axes, respectively, and by the definition of the coordinates of P, we have $\overrightarrow{P_1P} = \overrightarrow{OR} = z\hat{k}$. Similarly, $\overrightarrow{QP_1} = \overrightarrow{OS} = y\hat{j}$ and $\overrightarrow{OQ} = x\hat{i}$.

© NCERT not to be republished

Therefore, it follows that $\overrightarrow{OP_1} = \overrightarrow{OQ} + \overrightarrow{QP_1} = x\hat{i} + y\hat{j}$

and $\overrightarrow{OP} = \overrightarrow{OP_1} + \overrightarrow{P_1P} = x\hat{i} + y\hat{j} + z\hat{k}$

Hence, the position vector of P with reference to O is given by

$$\overrightarrow{OP} \text{ (or } \vec{r}) = x\hat{i} + y\hat{j} + z\hat{k}$$

This form of any vector is called its *component form.* Here, *x*, *y* and *z* are called as the *scalar components* of $\vec{r}$, and $x\hat{i}$, $y\hat{j}$ and $z\hat{k}$ are called the *vector components* of $\vec{r}$ along the respective axes. Sometimes *x*, *y* and *z* are also termed as *rectangular components*.

The length of any vector $\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$, is readily determined by applying the Pythagoras theorem twice. We note that in the right angle triangle OQP_1 (Fig 10.14)

$$|\overrightarrow{OP_1}| = \sqrt{|\overrightarrow{OQ}|^2 + |\overrightarrow{QP_1}|^2} = \sqrt{x^2 + y^2},$$

and in the right angle triangle OP_1P, we have

$$\overrightarrow{OP} = \sqrt{|\overrightarrow{OP_1}|^2 \quad |\overrightarrow{P_1P}|^2} \quad \sqrt{(x^2 \quad y^2) \quad z^2}$$

Hence, the length of any vector $\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$ is given by

$$|\vec{r}| = |x\hat{i} + y\hat{j} + z\hat{k}| = \sqrt{x^2 + y^2 + z^2}$$

If $\vec{a}$ and $\vec{b}$ are any two vectors given in the component form $a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$ and $b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$, respectively, then

(i) the sum (or resultant) of the vectors $\vec{a}$ and $\vec{b}$ is given by

$$\vec{a} + \vec{b} = (a_1 + b_1)\hat{i} + (a_2 + b_2)\hat{j} + (a_3 + b_3)\hat{k}$$

(ii) the difference of the vector $\vec{a}$ and $\vec{b}$ is given by

$$\vec{a} - \vec{b} = (a_1 - b_1)\hat{i} + (a_2 - b_2)\hat{j} + (a_3 - b_3)\hat{k}$$

(iii) the vectors $\vec{a}$ and $\vec{b}$ are equal if and only if

$$a_1 = b_1,\ a_2 = b_2 \quad \text{and} \quad a_3 = b_3$$

(iv) the multiplication of vector $\vec{a}$ by any scalar λ is given by

$$\lambda\vec{a} = (\ a_1)\hat{i} \quad (\ a_2)\hat{j} \quad (\ a_3)\hat{k}$$

© NCERT not to be republished

The addition of vectors and the multiplication of a vector by a scalar together give the following distributive laws:

Let $\vec{a}$ and $\vec{b}$ be any two vectors, and k and m be any scalars. Then

(i) $k\vec{a} + m\vec{a} = (k + m)\vec{a}$

(ii) $k(m\vec{a}) = (km)\vec{a}$

(iii) $k(\vec{a} \quad \vec{b}) \quad k\vec{a} \quad k\vec{b}$

Remarks

(i) One may observe that whatever be the value of λ, the vector $\lambda\vec{a}$ is always collinear to the vector $\vec{a}$. In fact, two vectors $\vec{a}$ and $\vec{b}$ are collinear if and only if there exists a nonzero scalar λ such that $\vec{b} = \lambda\vec{a}$. If the vectors $\vec{a}$ and $\vec{b}$ are given in the component form, i.e. $\vec{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$ and $\vec{b} = b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$, then the two vectors are collinear if and only if

$$b_1\hat{i} + b_2\hat{j} + b_3\hat{k} = \lambda(a_1\hat{i} + a_2\hat{j} + a_3\hat{k})$$

$$\Leftrightarrow \quad b_1\hat{i} + b_2\hat{j} + b_3\hat{k} = (\lambda a_1)\hat{i} + (\lambda a_2)\hat{j} + (\lambda a_3)\hat{k}$$

$$\Leftrightarrow \quad b_1 = \lambda a_1, \; b_2 = \lambda a_2, \; b_3 = \lambda a_3$$

$$\Leftrightarrow \quad \frac{b_1}{a_1} = \frac{b_2}{a_2} = \frac{b_3}{a_3} = \lambda$$

(ii) If $\vec{a} \quad a_1\hat{i} \quad a_2\hat{j} \quad a_3\hat{k}$, then a_1, a_2, a_3 are also called direction ratios of $\vec{a}$.

(iii) In case if it is given that l, m, n are direction cosines of a vector, then $l\hat{i} + m\hat{j} + n\hat{k}$ $= (\cos\alpha)\hat{i} + (\cos\beta)\hat{j} + (\cos\gamma)\hat{k}$ is the unit vector in the direction of that vector, where α, β and γ are the angles which the vector makes with x, y and z axes respectively.

Example 4 Find the values of x, y and z so that the vectors $\vec{a} = x\hat{i} + 2\hat{j} + z\hat{k}$ and $\vec{b} = 2\hat{i} + y\hat{j} + \hat{k}$ are equal.

Solution Note that two vectors are equal if and only if their corresponding components are equal. Thus, the given vectors $\vec{a}$ and $\vec{b}$ will be equal if and only if

$$x = 2, \; y = 2, \; z = 1$$

© NCERT not to be republished

Example 5 Let $\vec{a} = \hat{i} + 2\hat{j}$ and $\vec{b} = 2\hat{i} + \hat{j}$. Is $|\vec{a}| = |\vec{b}|$? Are the vectors $\vec{a}$ and $\vec{b}$ equal?

Solution We have $|\vec{a}| = \sqrt{1^2 + 2^2} = \sqrt{5}$ and $|\vec{b}|$ $\sqrt{2^2}$ 1^2 $\sqrt{5}$

So, $|\vec{a}| = |\vec{b}|$. But, the two vectors are not equal since their corresponding components are distinct.

Example 6 Find unit vector in the direction of vector $\vec{a} = 2\hat{i} + 3\hat{j} + \hat{k}$

Solution The unit vector in the direction of a vector $\vec{a}$ is given by $\hat{a} = \frac{1}{|\vec{a}|}\vec{a}$.

Now $$|\vec{a}| = \sqrt{2^2 + 3^2 + 1^2} = \sqrt{14}$$

Therefore $$\hat{a} = \frac{1}{\sqrt{14}}(2\hat{i} + 3\hat{j} + \hat{k}) = \frac{2}{\sqrt{14}}\hat{i} + \frac{3}{\sqrt{14}}\hat{j} + \frac{1}{\sqrt{14}}\hat{k}$$

Example 7 Find a vector in the direction of vector $\vec{a} = \hat{i} - 2\hat{j}$ that has magnitude 7 units.

Solution The unit vector in the direction of the given vector $\vec{a}$ is

$$\hat{a} = \frac{1}{|\vec{a}|}\vec{a} = \frac{1}{\sqrt{5}}(\hat{i} - 2\hat{j}) = \frac{1}{\sqrt{5}}\hat{i} - \frac{2}{\sqrt{5}}\hat{j}$$

Therefore, the vector having magnitude equal to 7 and in the direction of $\vec{a}$ is

$$7\hat{a} = 7\left(\frac{1}{\sqrt{5}}\hat{i} - \frac{2}{\sqrt{5}}\hat{j}\right) = \frac{7}{\sqrt{5}}\hat{i} - \frac{14}{\sqrt{5}}\hat{j}$$

Example 8 Find the unit vector in the direction of the sum of the vectors, $\vec{a} = 2\hat{i} + 2\hat{j} - 5\hat{k}$ and $\vec{b} = 2\hat{i} + \hat{j} + 3\hat{k}$.

Solution The sum of the given vectors is

$$\vec{a} \quad \vec{b}\,(\quad \vec{c}, \text{say}) = 4\hat{i} \quad 3\hat{j} \quad 2\hat{k}$$

and $$|\vec{c}| = \sqrt{4^2 + 3^2 + (-2)^2} = \sqrt{29}$$

© NCERT not to be republished

Thus, the required unit vector is

$$\hat{c} = \frac{1}{|\vec{c}|}\vec{c} = \frac{1}{\sqrt{29}}(4\hat{i} + 3\hat{j} - 2\hat{k}) = \frac{4}{\sqrt{29}}\hat{i} + \frac{3}{\sqrt{29}}\hat{j} - \frac{2}{\sqrt{29}}\hat{k}$$

Example 9 Write the direction ratio's of the vector $\vec{a} = \hat{i} + \hat{j} - 2\hat{k}$ and hence calculate its direction cosines.

Solution Note that the direction ratio's a, b, c of a vector $\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$ are just the respective components x, y and z of the vector. So, for the given vector, we have $a = 1$, $b = 1$ and $c = -2$. Further, if l, m and n are the direction cosines of the given vector, then

$$l = \frac{a}{|\vec{r}|} = \frac{1}{\sqrt{6}}, \quad m = \frac{b}{|\vec{r}|} = \frac{1}{\sqrt{6}}, \quad n = \frac{c}{|\vec{r}|} = \frac{-2}{\sqrt{6}} \text{ as } |\vec{r}| = \sqrt{6}$$

Thus, the direction cosines are $\left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right)$.

10.5.2 *Vector joining two points*

If $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ are any two points, then the vector joining P_1 and P_2 is the vector $\overrightarrow{P_1P_2}$ (Fig 10.15).

Joining the points P_1 and P_2 with the origin O, and applying triangle law, from the triangle OP_1P_2, we have

$$\overrightarrow{OP_1} + \overrightarrow{P_1P_2} = \overrightarrow{OP_2}.$$

Using the properties of vector addition, the above equation becomes

$$\overrightarrow{P_1P_2} = \overrightarrow{OP_2} - \overrightarrow{OP_1}$$

i.e.
$$\overrightarrow{P_1P_2} = (x_2\hat{i} + y_2\hat{j} + z_2\hat{k}) - (x_1\hat{i} + y_1\hat{j} + z_1\hat{k})$$
$$= (x_2 - x_1)\hat{i} + (y_2 - y_1)\hat{j} + (z_2 - z_1)\hat{k}$$

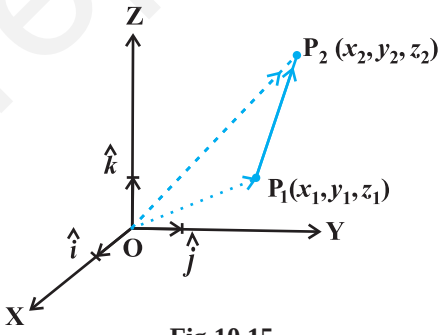


Fig 10.15

The magnitude of vector $\overrightarrow{P_1P_2}$ is given by

$$\overrightarrow{P_1P_2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

© NCERT not to be republished

Example 10 Find the vector joining the points P(2, 3, 0) and Q(– 1, – 2, – 4) directed from P to Q.

Solution Since the vector is to be directed from P to Q, clearly P is the initial point and Q is the terminal point. So, the required vector joining P and Q is the vector $\overrightarrow{PQ}$, given by

$$\overrightarrow{PQ} = (-1-2)\hat{i} + (-2-3)\hat{j} + (-4-0)\hat{k}$$

i.e. $$\overrightarrow{PQ} = -3\hat{i} - 5\hat{j} - 4\hat{k}.$$

10.5.3 *Section formula*

Let P and Q be two points represented by the position vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$, respectively, with respect to the origin O. Then the line segment joining the points P and Q may be divided by a third point, say R, in two ways – internally (Fig 10.16) and externally (Fig 10.17). Here, we intend to find the position vector $\overrightarrow{OR}$ for the point R with respect to the origin O. We take the two cases one by one.

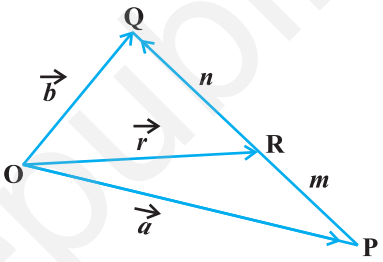


Fig 10.16

Case I When R divides PQ internally (Fig 10.16).

If R divides $\overrightarrow{PQ}$ such that $m\overrightarrow{RQ} = n\overrightarrow{PR}$,

where m and n are positive scalars, we say that the point R divides $\overrightarrow{PQ}$ internally in the ratio of $m : n$. Now from triangles ORQ and OPR, we have

$$\overrightarrow{RQ} = \overrightarrow{OQ} - \overrightarrow{OR} = \vec{b} - \vec{r}$$

and $$\overrightarrow{PR} = \overrightarrow{OR} - \overrightarrow{OP} = \vec{r} - \vec{a},$$

Therefore, we have $$m(\vec{b} - \vec{r}) = n(\vec{r} - \vec{a}) \quad \text{(Why?)}$$

or $$\vec{r} = \frac{m\vec{b} + n\vec{a}}{m+n} \quad \text{(on simplification)}$$

Hence, the position vector of the point R which divides P and Q internally in the ratio of $m : n$ is given by

$$\overrightarrow{OR} = \frac{m\vec{b} + n\vec{a}}{m+n}$$

© NCERT not to be republished

Case II When R divides PQ externally (Fig 10.17). We leave it to the reader as an exercise to verify that the position vector of the point R which divides the line segment PQ externally in the ratio $m : n$ i.e. $\frac{PR}{QR} \quad \frac{m}{n}$ is given by

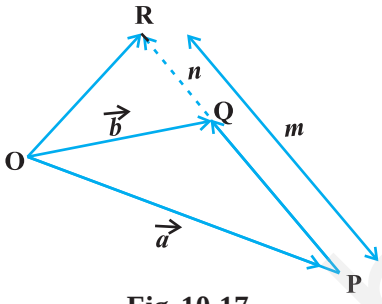


Fig 10.17

$$\overrightarrow{OR} = \frac{m\vec{b} - n\vec{a}}{m-n}$$

Remark If R is the midpoint of PQ , then $m = n$. And therefore, from Case I, the midpoint R of $\overrightarrow{PQ}$, will have its position vector as

$$\overrightarrow{OR} = \frac{\vec{a} + \vec{b}}{2}$$

Example 11 Consider two points P and Q with position vectors $\overrightarrow{OP} = 3\vec{a} - 2\vec{b}$ and $\overrightarrow{OQ} \quad \vec{a} \quad \vec{b}$. Find the position vector of a point R which divides the line joining P and Q in the ratio 2:1, (i) internally, and (ii) externally.

Solution

(i) The position vector of the point R dividing the join of P and Q internally in the ratio 2:1 is

$$\overrightarrow{OR} = \frac{2(\vec{a} + \vec{b}) + (3\vec{a} - 2\vec{b})}{2+1} = \frac{5\vec{a}}{3}$$

(ii) The position vector of the point R dividing the join of P and Q externally in the ratio 2:1 is

$$\overrightarrow{OR} = \frac{2(\vec{a} + \vec{b}) - (3\vec{a} - 2\vec{b})}{2-1} = 4\vec{b} - \vec{a}$$

Example 12 Show that the points $A(2\hat{i} \quad \hat{j} \quad \hat{k})$, $B(\hat{i} \quad 3\hat{j} \quad 5\hat{k})$, $C(3\hat{i} \quad 4j \quad 4\hat{k})$ are the vertices of a right angled triangle.

Solution We have

$$\overrightarrow{AB} = (1-2)\hat{i} + (-3+1)\hat{j} + (-5-1)\hat{k} \quad \hat{i} \quad 2\hat{j} \quad 6\hat{k}$$

$$\overrightarrow{BC} = (3-1)\hat{i} + (-4+3)\hat{j} + (-4+5)\hat{k} = 2\hat{i} - \hat{j} + \hat{k}$$

and $$\overrightarrow{CA} = (2-3)\hat{i} + (-1+4)\hat{j} + (1+4)\hat{k} = -\hat{i} + 3\hat{j} + 5\hat{k}$$

© NCERT not to be republished

Further, note that

$$|\overrightarrow{AB}|^2 = 41 = 6 + 35 = |\overrightarrow{BC}|^2 + |\overrightarrow{CA}|^2$$

Hence, the triangle is a right angled triangle.

EXERCISE 10.2

1. Compute the magnitude of the following vectors:

$$\vec{a} = \hat{i} + \hat{j} + k; \quad \vec{b} = 2\hat{i} - 7\hat{j} - 3\hat{k}; \quad \vec{c} = \frac{1}{\sqrt{3}}\hat{i} + \frac{1}{\sqrt{3}}\hat{j} - \frac{1}{\sqrt{3}}\hat{k}$$

2. Write two different vectors having same magnitude.

3. Write two different vectors having same direction.

4. Find the values of x and y so that the vectors $2\hat{i} + 3\hat{j}$ and $x\hat{i} + y\hat{j}$ are equal.

5. Find the scalar and vector components of the vector with initial point (2, 1) and terminal point (– 5, 7).

6. Find the sum of the vectors $\vec{a} = \hat{i} - 2\hat{j} + \hat{k}$, $\vec{b} = -2\hat{i} + 4\hat{j} + 5\hat{k}$ and $\vec{c} = \hat{i} - 6\hat{j} - 7\hat{k}$.

7. Find the unit vector in the direction of the vector $\vec{a} = \hat{i} + \hat{j} + 2\hat{k}$.

8. Find the unit vector in the direction of vector $\overrightarrow{PQ}$, where P and Q are the points (1, 2, 3) and (4, 5, 6), respectively.

9. For given vectors, $\vec{a} = 2\hat{i} - \hat{j} + 2\hat{k}$ and $\vec{b} = -\hat{i} + \hat{j} - \hat{k}$, find the unit vector in the direction of the vector $\vec{a} + \vec{b}$.

10. Find a vector in the direction of vector $5\hat{i} - \hat{j} + 2\hat{k}$ which has magnitude 8 units.

11. Show that the vectors $2\hat{i} - 3\hat{j} + 4\hat{k}$ and $-4\hat{i} + 6\hat{j} - 8\hat{k}$ are collinear.

12. Find the direction cosines of the vector $\hat{i} + 2\hat{j} + 3\hat{k}$.

13. Find the direction cosines of the vector joining the points A (1, 2, –3) and B(–1, –2, 1), directed from A to B.

14. Show that the vector $\hat{i} + \hat{j} + \hat{k}$ is equally inclined to the axes OX, OY and OZ.

15. Find the position vector of a point R which divides the line joining two points P and Q whose position vectors are $\hat{i} + 2\hat{j} - \hat{k}$ and $-\hat{i} + \hat{j} + \hat{k}$ respectively, in the ratio 2 : 1

(i) internally (ii) externally

© NCERT not to be republished

16. Find the position vector of the mid point of the vector joining the points P(2, 3, 4) and Q(4, 1, –2).

17. Show that the points A, B and C with position vectors, $\vec{a} = 3\hat{i} - 4\hat{j} - 4\hat{k}$, $\vec{b} = 2\hat{i} - \hat{j} + \hat{k}$ and $\vec{c} = \hat{i} - 3\hat{j} - 5\hat{k}$, respectively form the vertices of a right angled triangle.

18. In triangle ABC (Fig 10.18), which of the following is not true:

(A) $\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA} = \vec{0}$

(B) $\overrightarrow{AB} + \overrightarrow{BC} - \overrightarrow{AC} = \vec{0}$

(C) $\overrightarrow{AB} + \overrightarrow{BC} - \overrightarrow{CA} = \vec{0}$

(D) $\overrightarrow{AB} - \overrightarrow{CB} + \overrightarrow{CA} = \vec{0}$

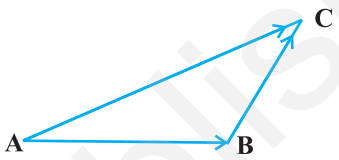


Fig 10.18

19. If $\vec{a}$ and $\vec{b}$ are two collinear vectors, then which of the following are incorrect:

(A) $\vec{b} = \lambda\vec{a}$, for some scalar λ

(B) $\vec{a} = \pm\vec{b}$

(C) the respective components of $\vec{a}$ and $\vec{b}$ are not proportional

(D) both the vectors $\vec{a}$ and $\vec{b}$ have same direction, but different magnitudes.

10.6 Product of Two Vectors

So far we have studied about addition and subtraction of vectors. An other algebraic operation which we intend to discuss regarding vectors is their product. We may recall that product of two numbers is a number, product of two matrices is again a matrix. But in case of functions, we may multiply them in two ways, namely, multiplication of two functions pointwise and composition of two functions. Similarly, multiplication of two vectors is also defined in two ways, namely, scalar (or dot) product where the result is a scalar, and vector (or cross) product where the result is a vector. Based upon these two types of products for vectors, they have found various applications in geometry, mechanics and engineering. In this section, we will discuss these two types of products.

10.6.1 *Scalar (or dot) product of two vectors*

Definition 2 The scalar product of two nonzero vectors $\vec{a}$ and $\vec{b}$, denoted by $\vec{a} \cdot \vec{b}$, is

© NCERT not to be republished

defined as $\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}|\cos\theta,$

where, θ is the angle between $\vec{a}$ and $\vec{b}$, 0 (Fig 10.19).

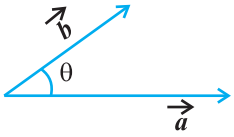


Fig 10.19

If either $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$, then θ is not defined, and in this case, we define $\vec{a}$ $\vec{b}$ 0

Observations

1. $\vec{a} \cdot \vec{b}$ is a real number.
2. Let $\vec{a}$ and $\vec{b}$ be two nonzero vectors, then $\vec{a} \cdot \vec{b} = 0$ if and only if $\vec{a}$ and $\vec{b}$ are perpendicular to each other. i.e.

   $\vec{a} \cdot \vec{b} = 0 \Leftrightarrow \vec{a} \perp \vec{b}$
3. If θ = 0, then $\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}|$

   In particular, $\vec{a} \cdot \vec{a} = |\vec{a}|^2$, as θ in this case is 0.
4. If θ = π, then $\vec{a} \cdot \vec{b} = -|\vec{a}||\vec{b}|$

   In particular, $\vec{a}$ ($\vec{a}$) $|\vec{a}|^2$, as θ in this case is π.
5. In view of the Observations 2 and 3, for mutually perpendicular unit vectors $\hat{i}$, $\hat{j}$ and $\hat{k}$, we have

   $$\hat{i} \cdot \hat{i} = \hat{j} \cdot \hat{j} = \hat{k} \cdot \hat{k} = 1,$$

   $$\hat{i} \cdot \hat{j} = \hat{j} \cdot \hat{k} = \hat{k} \ \hat{i} \quad 0$$
6. The angle between two nonzero vectors $\vec{a}$ and $\vec{b}$ is given by

   $$\cos \frac{\vec{a} \ \vec{b}}{|\vec{a}||\vec{b}|}, \text{ or } \theta = \cos^{-1}\left(\frac{\vec{a}.\vec{b}}{|\vec{a}||\vec{b}|}\right)$$
7. The scalar product is commutative. i.e.

   $$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$$ (Why?)

Two important properties of scalar product

Property 1 (Distributivity of scalar product over addition) Let $\vec{a}$, $\vec{b}$ and $\vec{c}$ be any three vectors, then

$$\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$$

© NCERT not to be republished

Property 2 Let $\vec{a}$ and $\vec{b}$ be any two vectors, and λ be any scalar. Then

$$(\lambda\vec{a})\cdot\vec{b} = (\vec{a})\ \vec{b} \quad (\vec{a}\ \vec{b}) \quad \vec{a}\ (\vec{b})$$

If two vectors $\vec{a}$ and $\vec{b}$ are given in component form as $a_1\hat{i}+a_2\hat{j}+a_3\hat{k}$ and $b_1\hat{i}+b_2\hat{j}+b_3\hat{k}$, then their scalar product is given as

$$\begin{aligned}\vec{a}\cdot\vec{b} &= (a_1\hat{i}+a_2\hat{j}+a_3\hat{k})\cdot(b_1\hat{i}+b_2\hat{j}+b_3\hat{k})\\ &= a_1\hat{i}\cdot(b_1\hat{i}+b_2\hat{j}+b_3\hat{k})+a_2\hat{j}\cdot(b_1\hat{i}+b_2\hat{j}+b_3\hat{k})+a_3\hat{k}\cdot(b_1\hat{i}+b_2\hat{j}+b_3\hat{k})\\ &= a_1b_1(\hat{i}\cdot\hat{i})+a_1b_2(\hat{i}\cdot\hat{j})+a_1b_3(\hat{i}\cdot\hat{k})+a_2b_1(\hat{j}\cdot\hat{i})+a_2b_2(\hat{j}\cdot\hat{j})+a_2b_3(\hat{j}\cdot\hat{k})\\ &\quad + a_3b_1(\hat{k}\cdot\hat{i})+a_3b_2(\hat{k}\cdot\hat{j})+a_3b_3(\hat{k}\cdot\hat{k}) \text{ (Using the above Properties 1 and 2)}\\ &= a_1b_1 + a_2b_2 + a_3b_3 \qquad \text{(Using Observation 5)}\end{aligned}$$

Thus $\vec{a}\cdot\vec{b} = a_1b_1+a_2b_2+a_3b_3$

10.6.2 *Projection of a vector on a line*

Suppose a vector $\overrightarrow{AB}$ makes an angle θ with a given directed line *l* (say), in the *anticlockwise direction* (Fig 10.20). Then the projection of $\overrightarrow{AB}$ on *l* is a vector $\vec{p}$ (say) with magnitude $|\overrightarrow{AB}|\cos\theta$, and the direction of $\vec{p}$ being the same (or opposite) to that of the line *l*, depending upon whether cos θ is positive or negative. The vector $\vec{p}$

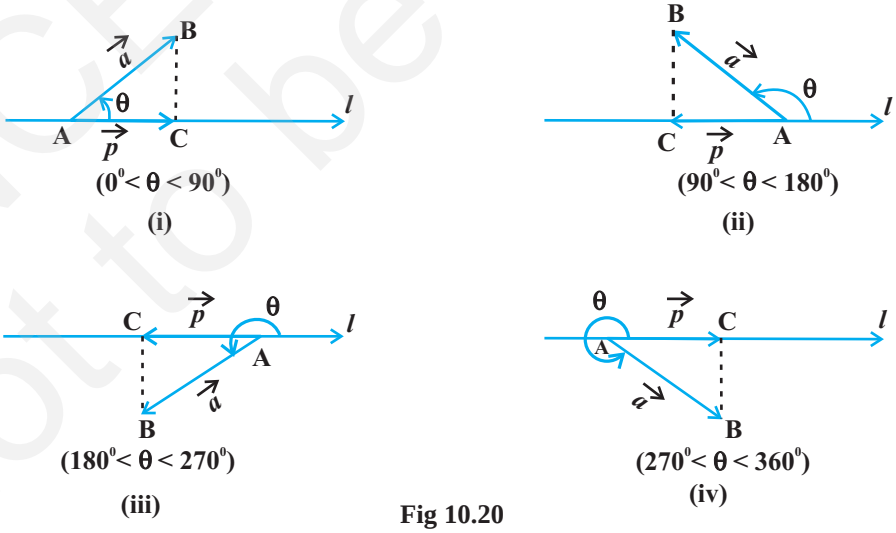


Fig 10.20

© NCERT not to be republished

is called the *projection vector*, and its magnitude $|\vec{p}|$ is simply called as the *projection* of the vector $\overrightarrow{AB}$ on the directed line *l*.

For example, in each of the following figures (Fig 10.20 (i) to (iv)), projection vector of $\overrightarrow{AB}$ along the line *l* is vector $\overrightarrow{AC}$.

Observations

1. If $\hat{p}$ is the unit vector along a line *l*, then the projection of a vector $\vec{a}$ on the line *l* is given by $\vec{a}\ \hat{p}$.
2. Projection of a vector $\vec{a}$ on other vector $\vec{b}$, is given by

$$\vec{a}\cdot\hat{b}, \quad \text{or} \quad \vec{a}\cdot\left(\frac{\vec{b}}{|\vec{b}|}\right), \quad \text{or} \quad \frac{1}{|\vec{b}|}(\vec{a}\cdot\vec{b})$$

3. If $\theta = 0$, then the projection vector of $\overrightarrow{AB}$ will be $\overrightarrow{AB}$ itself and if $\theta = \pi$, then the projection vector of $\overrightarrow{AB}$ will be $\overrightarrow{BA}$.
4. If $\theta = \frac{\pi}{2}$ or $\theta = \frac{3\pi}{2}$, then the projection vector of $\overrightarrow{AB}$ will be zero vector.

Remark If α, β and γ are the direction angles of vector $\vec{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$, then its direction cosines may be given as

$$\cos \quad \frac{\vec{a}\ \hat{i}}{|\vec{a}||\hat{i}|} \quad \frac{a_1}{|\vec{a}|}, \quad \cos \quad \frac{a_2}{|\vec{a}|}, \quad \text{and} \quad \cos \quad \frac{a_3}{|\vec{a}|}$$

Also, note that $|\vec{a}|\cos\alpha$, $|\vec{a}|\cos\beta$ and $|\vec{a}|\cos\gamma$ are respectively the projections of $\vec{a}$ along OX, OY and OZ. i.e., the scalar components a_1, a_2 and a_3 of the vector $\vec{a}$, are precisely the projections of $\vec{a}$ along *x*-axis, *y*-axis and *z*-axis, respectively. Further, if $\vec{a}$ is a unit vector, then it may be expressed in terms of its direction cosines as

$$\vec{a} = \cos\alpha\hat{i} + \cos\beta\hat{j} + \cos\gamma\hat{k}$$

Example 13 Find the angle between two vectors $\vec{a}$ and $\vec{b}$ with magnitudes 1 and 2 respectively and when $\vec{a}\cdot\vec{b} = 1$.

Solution Given $\vec{a}\ \vec{b}\quad 1, |\vec{a}|\quad 1$ and $|\vec{b}|\quad 2$. We have

$$\cos^{1}\ \frac{\vec{a}\ \vec{b}}{|\vec{a}||\vec{b}|}\quad \cos^{1}\ \frac{1}{2}\quad \frac{}{3}$$

© NCERT not to be republished

Example 14 Find angle 'θ' between the vectors $\vec{a} = \hat{i} + \hat{j} - \hat{k}$ and $\vec{b} = \hat{i} - \hat{j} + \hat{k}$.

Solution The angle θ between two vectors $\vec{a}$ and $\vec{b}$ is given by

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|}$$

Now $\vec{a} \cdot \vec{b} = (\hat{i} + \hat{j} - \hat{k}) \cdot (\hat{i} - \hat{j} + \hat{k}) = 1 - 1 - 1 = -1.$

Therefore, we have $\cos\theta = \frac{-1}{3}$

hence the required angle is $\theta = \cos^{-1}\frac{1}{3}$

Example 15 If $\vec{a} = 5\hat{i} - \hat{j} - 3\hat{k}$ and $\vec{b} = \hat{i} + 3\hat{j} - 5\hat{k}$, then show that the vectors $\vec{a} + \vec{b}$ and $\vec{a} - \vec{b}$ are perpendicular.

Solution We know that two nonzero vectors are perpendicular if their scalar product is zero.

Here $\vec{a} + \vec{b} = (5\hat{i} - \hat{j} - 3\hat{k}) + (\hat{i} + 3\hat{j} - 5\hat{k}) = 6\hat{i} + 2\hat{j} - 8\hat{k}$

and $\vec{a} - \vec{b} = (5\hat{i} - \hat{j} - 3\hat{k}) - (\hat{i} + 3\hat{j} - 5\hat{k}) = 4\hat{i} - 4\hat{j} + 2\hat{k}$

So $(\vec{a} + \vec{b}) \cdot (\vec{a} - \vec{b}) = (6\hat{i} + 2\hat{j} - 8\hat{k}) \cdot (4\hat{i} - 4\hat{j} + 2\hat{k}) = 24 - 8 - 16 = 0.$

Hence $\vec{a} + \vec{b}$ and $\vec{a} - \vec{b}$ are perpendicular vectors.

Example 16 Find the projection of the vector $\vec{a} = 2\hat{i} + 3\hat{j} + 2\hat{k}$ on the vector $\vec{b} = \hat{i} + 2\hat{j} + \hat{k}$.

Solution The projection of vector $\vec{a}$ on the vector $\vec{b}$ is given by

$$\frac{1}{|\vec{b}|}(\vec{a} \cdot \vec{b}) = \frac{(2 \times 1 + 3 \times 2 + 2 \times 1)}{\sqrt{(1)^2 + (2)^2 + (1)^2}} = \frac{10}{\sqrt{6}} = \frac{5}{3}\sqrt{6}$$

Example 17 Find $|\vec{a} - \vec{b}|$, if two vectors $\vec{a}$ and $\vec{b}$ are such that $|\vec{a}| \quad 2$, $|\vec{b}| \quad 3$ and $\vec{a} \cdot \vec{b} = 4$.

Solution We have

$$|\vec{a} \quad \vec{b}|^2 = (\vec{a} - \vec{b}) \cdot (\vec{a} - \vec{b})$$
$$= \vec{a}.\vec{a} - \vec{a} \cdot \vec{b} - \vec{b} \cdot \vec{a} + \vec{b} \cdot \vec{b}$$

© NCERT not to be republished

$$= |\vec{a}|^2 - 2(\vec{a}\cdot\vec{b}) + |\vec{b}|^2$$

$$= (2)^2 - 2(4) + (3)^2$$

Therefore $|\vec{a} - \vec{b}| = \sqrt{5}$

Example 18 If $\vec{a}$ is a unit vector and $(\vec{x} - \vec{a})\cdot(\vec{x} + \vec{a}) = 8$, then find $|\vec{x}|$.

Solution Since $\vec{a}$ is a unit vector, $|\vec{a}| = 1$. Also,

$$(\vec{x} - \vec{a})\cdot(\vec{x} + \vec{a}) = 8$$

or $$\vec{x}\cdot\vec{x} + \vec{x}\cdot\vec{a} - \vec{a}\cdot\vec{x} - \vec{a}\cdot\vec{a} = 8$$

or $$|\vec{x}|^2 \quad 1 = 8 \text{ i.e. } |\vec{x}|^2 = 9$$

Therefore $|\vec{x}| = 3$ (as magnitude of a vector is non negative).

Example 19 For any two vectors $\vec{a}$ and $\vec{b}$, we always have $|\vec{a}\cdot\vec{b}| \le |\vec{a}||\vec{b}|$ (Cauchy-Schwartz inequality).

Solution The inequality holds trivially when either $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$. Actually, in such a situation we have $|\vec{a}\cdot\vec{b}| = 0 = |\vec{a}||\vec{b}|$. So, let us assume that $|\vec{a}| \neq 0 \neq |\vec{b}|$. Then, we have

$$\frac{|\vec{a}\cdot\vec{b}|}{|\vec{a}||\vec{b}|} = |\cos\theta| \le 1$$

Therefore $|\vec{a}\cdot\vec{b}| \le |\vec{a}||\vec{b}|$

Example 20 For any two vectors $\vec{a}$ and $\vec{b}$, we always have $|\vec{a} + \vec{b}| \le |\vec{a}| + |\vec{b}|$ (triangle inequality).

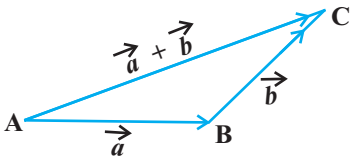


Fig 10.21

Solution The inequality holds trivially in case either $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$ (How?). So, let $|\vec{a}| \quad \vec{0} \quad |\vec{b}|$. Then,

$$|\vec{a} + \vec{b}|^2 = (\vec{a} + \vec{b})^2 = (\vec{a} + \vec{b})\cdot(\vec{a} + \vec{b})$$

$$= \vec{a}\cdot\vec{a} + \vec{a}\cdot\vec{b} + \vec{b}\cdot\vec{a} + \vec{b}\cdot\vec{b}$$

$$= |\vec{a}|^2 + 2\vec{a}\cdot\vec{b} + |\vec{b}|^2 \quad \text{(scalar product is commutative)}$$

$$\le |\vec{a}|^2 + 2|\vec{a}\cdot\vec{b}| + |\vec{b}|^2 \quad \text{(since } x \le |x| \ \forall x \in \mathbf{R}\text{)}$$

$$\le |\vec{a}|^2 + 2|\vec{a}||\vec{b}| + |\vec{b}|^2 \quad \text{(from Example 19)}$$

$$= (|\vec{a}| \quad |\vec{b}|)^2$$

© NCERT not to be republished

Hence $|\vec{a} \quad \vec{b}| \leq |\vec{a}| \quad |\vec{b}|$

Remark If the equality holds in triangle inequality (in the above Example 20), i.e.

$$|\vec{a}+\vec{b}| = |\vec{a}|+|\vec{b}|,$$

then $$|\overrightarrow{AC}| = |\overrightarrow{AB}|+|\overrightarrow{BC}|$$

showing that the points A, B and C are collinear.

Example 21 Show that the points $A(-2\hat{i}+3\hat{j}+5\hat{k})$, $B(\hat{i}+2\hat{j}+3\hat{k})$ and $C(7\hat{i}-\hat{k})$ are collinear.

Solution We have

$$\overrightarrow{AB} = (1 \quad 2)\hat{i} \quad (2 \quad 3)\hat{j} \quad (3 \quad 5)\hat{k} \quad 3\hat{i} \quad \hat{j} \quad 2\hat{k},$$

$$\overrightarrow{BC} = (7 \quad 1)\hat{i} \quad (0 \quad 2)\hat{j} \quad (\ 1 \quad 3)\hat{k} \quad 6\hat{i} \quad 2\hat{j} \quad 4\hat{k},$$

$$\overrightarrow{AC} = (7 \quad 2)\hat{i} \quad (0 \quad 3)\hat{j} \quad (\ 1 \quad 5)\hat{k} \quad 9\hat{i} \quad 3\hat{j} \quad 6\hat{k}$$

$$|\overrightarrow{AB}| = \sqrt{14}, \ |\overrightarrow{BC}| \quad 2\sqrt{14} \text{ and } |\overrightarrow{AC}| \quad 3\sqrt{14}$$

Therefore $$\left|\overrightarrow{AC}\right| = |\overrightarrow{AB}|+|\overrightarrow{BC}|$$

Hence the points A, B and C are collinear.

Note In Example 21, one may note that although $\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CA}=\vec{0}$ but the points A, B and C do not form the vertices of a triangle.

EXERCISE 10.3

1. Find the angle between two vectors $\vec{a}$ and $\vec{b}$ with magnitudes $\sqrt{3}$ and 2, respectively having $\vec{a}\cdot\vec{b}=\sqrt{6}$.
2. Find the angle between the vectors $\hat{i}-2\hat{j}+3\hat{k}$ and $3\hat{i}-2\hat{j}+\hat{k}$
3. Find the projection of the vector $\hat{i}-\hat{j}$ on the vector $\hat{i}+\hat{j}$.
4. Find the projection of the vector $\hat{i}+3\hat{j}+7\hat{k}$ on the vector $7\hat{i}-\hat{j}+8\hat{k}$.
5. Show that each of the given three vectors is a unit vector:

   $\frac{1}{7}(2\hat{i}+3\hat{j}+6\hat{k}), \ \frac{1}{7}(3\hat{i}-6\hat{j}+2\hat{k}), \ \frac{1}{7}(6\hat{i}+2\hat{j}-3\hat{k})$

   Also, show that they are mutually perpendicular to each other.

© NCERT not to be republished

6. Find $|\vec{a}|$ and $|\vec{b}|$, if $(\vec{a}+\vec{b})\cdot(\vec{a}-\vec{b})=8$ and $|\vec{a}|=8|\vec{b}|$.

7. Evaluate the product $(3\vec{a}-5\vec{b})\cdot(2\vec{a}+7\vec{b})$.

8. Find the magnitude of two vectors $\vec{a}$ and $\vec{b}$, having the same magnitude and such that the angle between them is 60^o and their scalar product is $\frac{1}{2}$.

9. Find $|\vec{x}|$, if for a unit vector $\vec{a}$, $(\vec{x}-\vec{a})\cdot(\vec{x}+\vec{a})=12$.

10. If $\vec{a}=2\hat{i}+2\hat{j}+3\hat{k}$, $\vec{b}=-\hat{i}+2\hat{j}+\hat{k}$ and $\vec{c}=3\hat{i}+\hat{j}$ are such that $\vec{a}+\lambda\vec{b}$ is perpendicular to $\vec{c}$, then find the value of λ.

11. Show that $|\vec{a}|\vec{b}+|\vec{b}|\vec{a}$ is perpendicular to $|\vec{a}|\vec{b}-|\vec{b}|\vec{a}$, for any two nonzero vectors $\vec{a}$ and $\vec{b}$.

12. If $\vec{a}\cdot\vec{a}=0$ and $\vec{a}\cdot\vec{b}=0$, then what can be concluded about the vector $\vec{b}$?

13. If $\vec{a},\vec{b},\vec{c}$ are unit vectors such that $\vec{a}+\vec{b}+\vec{c}=\vec{0}$, find the value of $\vec{a}\cdot\vec{b}+\vec{b}\cdot\vec{c}+\vec{c}\cdot\vec{a}$.

14. If either vector $\vec{a}=\vec{0}$ or $\vec{b}=\vec{0}$, then $\vec{a}\cdot\vec{b}=0$. But the converse need not be true. Justify your answer with an example.

15. If the vertices A, B, C of a triangle ABC are (1, 2, 3), (–1, 0, 0), (0, 1, 2), respectively, then find ∠ABC. [∠ABC is the angle between the vectors $\overrightarrow{BA}$ and $\overrightarrow{BC}$].

16. Show that the points A(1, 2, 7), B(2, 6, 3) and C(3, 10, –1) are collinear.

17. Show that the vectors $2\hat{i}-\hat{j}+\hat{k}$, $\hat{i}-3\hat{j}-5\hat{k}$ and $3\hat{i}-4\hat{j}-4\hat{k}$ form the vertices of a right angled triangle.

18. If $\vec{a}$ is a nonzero vector of magnitude '*a*' and λ a nonzero scalar, then $\lambda\vec{a}$ is unit vector if

(A) $\lambda = 1$ (B) $\lambda = -1$ (C) $a = |\lambda|$ (D) $a = 1/|\lambda|$

10.6.3 *Vector (or cross) product of two vectors*

In Section 10.2, we have discussed on the three dimensional right handed rectangular coordinate system. In this system, when the positive *x*-axis is rotated counterclockwise

© NCERT not to be republished

into the positive *y*-axis, a right handed (standard) screw would advance in the direction of the positive *z*-axis (Fig 10.22(i)).

In a right handed coordinate system, the thumb of the right hand points in the direction of the positive *z*-axis when the fingers are curled in the direction away from the positive *x*-axis toward the positive *y*-axis (Fig 10.22(ii)).

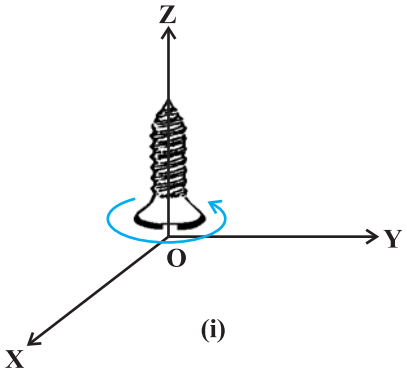


Fig 10.22 (i), (ii)

Definition 3 The vector product of two nonzero vectors $\vec{a}$ and $\vec{b}$, is denoted by $\vec{a} \quad \vec{b}$ and defined as

$$\vec{a} \times \vec{b} = |\vec{a}||\vec{b}| \sin\theta \, \hat{n},$$

where, θ is the angle between $\vec{a}$ and $\vec{b}$, $0 \le \theta \le \pi$ and $\hat{n}$ is a unit vector perpendicular to both $\vec{a}$ and $\vec{b}$, such that $\vec{a}, \vec{b}$ and $\hat{n}$ form a right handed system (Fig 10.23). i.e., the right handed system rotated from $\vec{a}$ to $\vec{b}$ moves in the direction of $\hat{n}$.

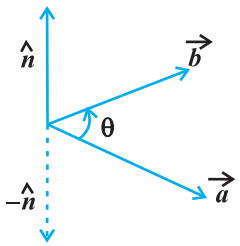


Fig 10.23

If either $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$, then θ is not defined and in this case, we define $\vec{a} \times \vec{b} = \vec{0}$.

Observations

1. $\vec{a} \times \vec{b}$ is a vector.
2. Let $\vec{a}$ and $\vec{b}$ be two nonzero vectors. Then $\vec{a} \times \vec{b} = \vec{0}$ if and only if $\vec{a}$ and $\vec{b}$ are parallel (or collinear) to each other, i.e.,

$$\vec{a} \times \vec{b} = \vec{0} \Leftrightarrow \vec{a} \parallel \vec{b}$$

© NCERT not to be republished

In particular, $\vec{a} \times \vec{a} = \vec{0}$ and $\vec{a} \times (-\vec{a}) = \vec{0}$, since in the first situation, $\theta = 0$ and in the second one, $\theta = \pi$, making the value of $\sin\theta$ to be 0.

3. If $\quad$ then $\vec{a} \quad \vec{b} \quad |\vec{a}||\vec{b}|$.

4. In view of the Observations 2 and 3, for mutually perpendicular unit vectors $\hat{i}$, $\hat{j}$ and $\hat{k}$ (Fig 10.24), we have

$$\hat{i} \times \hat{i} = \hat{j} \times \hat{j} = \hat{k} \times \hat{k} = \vec{0}$$

$$\hat{i} \times \hat{j} = \hat{k}, \quad \hat{j} \times \hat{k} = \hat{i}, \quad \hat{k} \times \hat{i} = \hat{j}$$

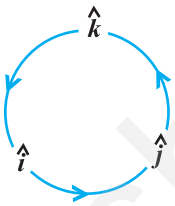


Fig 10.24

5. In terms of vector product, the angle between two vectors $\vec{a}$ and $\vec{b}$ may be given as

$$\sin\theta = \frac{|\vec{a} \times \vec{b}|}{|\vec{a}||\vec{b}|}$$

6. It is always true that the vector product is not commutative, as $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$. Indeed, $\vec{a} \times \vec{b} = |\vec{a}||\vec{b}| \sin\theta \hat{n}$, where $\vec{a}, \vec{b}$ and $\hat{n}$ form a right handed system, i.e., θ is traversed from $\vec{a}$ to $\vec{b}$, Fig 10.25 (i). While, $\vec{b} \times \vec{a} = |\vec{a}||\vec{b}| \sin\theta \hat{n}_1$, where $\vec{b}, \vec{a}$ and $\hat{n}_1$ form a right handed system i.e. θ is traversed from $\vec{b}$ to $\vec{a}$, Fig 10.25(ii).

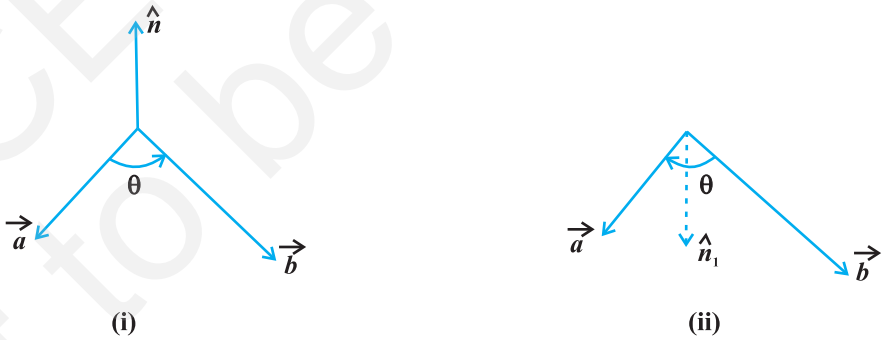


Fig 10.25 (i), (ii)

Thus, if we assume $\vec{a}$ and $\vec{b}$ to lie in the plane of the paper, then $\hat{n}$ and $\hat{n}_1$ both will be perpendicular to the plane of the paper. But, $\hat{n}$ being directed above the paper while $\hat{n}_1$ directed below the paper. i.e. $\hat{n}_1 = -\hat{n}$.

© NCERT not to be republished

Hence $\vec{a} \times \vec{b} = |\vec{a}||\vec{b}| \sin \hat{n}$

$= -|\vec{a}||\vec{b}| \sin\theta \hat{n}_1 = -\vec{b} \times \vec{a}$

7. In view of the Observations 4 and 6, we have

$$\hat{j} \times \hat{i} = -\hat{k}, \quad \hat{k} \times \hat{j} = -\hat{i} \text{ and } \hat{i} \times \hat{k} = -\hat{j}.$$

8. If $\vec{a}$ and $\vec{b}$ represent the adjacent sides of a triangle then its area is given as $\frac{1}{2}|\vec{a} \quad \vec{b}|$.

By definition of the area of a triangle, we have from Fig 10.26,

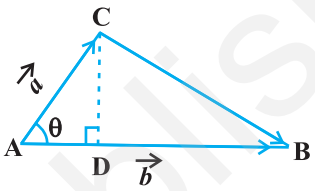


Fig 10.26

Area of triangle ABC = $\frac{1}{2}$ AB · CD.

But AB = $|\vec{b}|$ (as given), and CD = $|\vec{a}| \sin\theta$.

Thus, Area of triangle ABC = $\frac{1}{2}|\vec{b}||\vec{a}| \sin\theta = \frac{1}{2}|\vec{a} \times \vec{b}|$.

9. If $\vec{a}$ and $\vec{b}$ represent the adjacent sides of a parallelogram, then its area is given by $|\vec{a} \times \vec{b}|$.

From Fig 10.27, we have

Area of parallelogram ABCD = AB. DE.

But AB = $|\vec{b}|$ (as given), and

DE = $|\vec{a}| \sin\theta$.

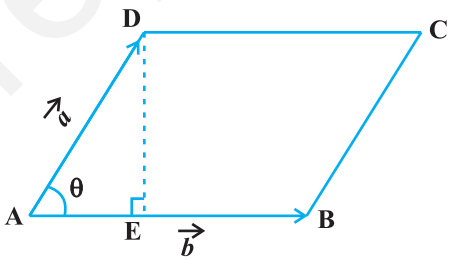


Fig 10.27

Thus,

Area of parallelogram ABCD = $|\vec{b}||\vec{a}| \sin\theta = |\vec{a} \times \vec{b}|$.

We now state two important properties of vector product.

Property 3 (Distributivity of vector product over addition): If $\vec{a}$, $\vec{b}$ and $\vec{c}$ are any three vectors and λ be a scalar, then

(i) $\vec{a} \times (\vec{b} + \vec{c}) = \vec{a} \quad \vec{b} \quad \vec{a} \quad \vec{c}$

(ii) $\lambda(\vec{a} \times \vec{b}) = (\lambda\vec{a}) \times \vec{b} = \vec{a} \times (\lambda\vec{b})$

© NCERT not to be republished

Let $\vec{a}$ and $\vec{b}$ be two vectors given in component form as $a_1\hat{i}+a_2\hat{j}+a_3\hat{k}$ and $b_1\hat{i}+b_2\hat{j}+b_3\hat{k}$, respectively. Then their cross product may be given by

$$\vec{a}\times\vec{b} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Explanation We have

$$\begin{aligned}
\vec{a}\times\vec{b} &= (a_1\hat{i}+a_2\hat{j}+a_3\hat{k})\times(b_1\hat{i}+b_2\hat{j}+b_3\hat{k}) \\
&= a_1b_1(\hat{i}\times\hat{i})+a_1b_2(\hat{i}\times\hat{j})+a_1b_3(\hat{i}\times\hat{k})+a_2b_1(\hat{j}\times\hat{i}) \\
&\quad + a_2b_2(\hat{j}\times\hat{j})+a_2b_3(\hat{j}\times\hat{k}) \\
&\quad + a_3b_1(\hat{k}\times\hat{i})+a_3b_2(\hat{k}\times\hat{j})+a_3b_3(\hat{k}\times\hat{k}) \qquad \text{(by Property 1)} \\
&= a_1b_2(\hat{i}\times\hat{j})-a_1b_3(\hat{k}\times\hat{i})-a_2b_1(\hat{i}\times\hat{j}) \\
&\quad + a_2b_3(\hat{j}\times\hat{k})+a_3b_1(\hat{k}\times\hat{i})-a_3b_2(\hat{j}\times\hat{k})
\end{aligned}$$

(as $\hat{i}\times\hat{i}=\hat{j}\times\hat{j}=\hat{k}\times\hat{k}=0$ and $\hat{i}\times\hat{k}=-\hat{k}\times\hat{i}$, $\hat{j}\times\hat{i}=-\hat{i}\times\hat{j}$ and $\hat{k}\times\hat{j}=-\hat{j}\times\hat{k}$)

$$\begin{aligned}
&= a_1b_2\hat{k}-a_1b_3\hat{j}-a_2b_1\hat{k}+a_2b_3\hat{i}+a_3b_1\hat{j}-a_3b_2\hat{i} \\
&\quad (\text{as } \hat{i}\times\hat{j}=\hat{k},\ \hat{j}\times\hat{k}=\hat{i} \text{ and } \hat{k}\times\hat{i}=\hat{j}) \\
&= (a_2b_3-a_3b_2)\hat{i}-(a_1b_3-a_3b_1)\hat{j}+(a_1b_2-a_2b_1)\hat{k} \\
&= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}
\end{aligned}$$

Example 22 Find $|\vec{a}\times\vec{b}|$, if $\vec{a}=2\hat{i}+\hat{j}+3\hat{k}$ and $\vec{b}=3\hat{i}+5\hat{j}-2\hat{k}$

Solution We have

$$\vec{a}\times\vec{b} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 2 & 1 & 3 \\ 3 & 5 & -2 \end{vmatrix}$$

$$= \hat{i}(-2-15)-(-4-9)\hat{j}+(10-3)\hat{k} = -17\hat{i}+13\hat{j}+7\hat{k}$$

Hence $|\vec{a}\ \ \vec{b}| = \sqrt{(-17)^2+(13)^2+(7)^2} = \sqrt{507}$

© NCERT not to be republished

Example 23 Find a unit vector perpendicular to each of the vectors $(\vec{a}+\vec{b})$ and $(\vec{a}-\vec{b})$, where $\vec{a}=\hat{i}+\hat{j}+\hat{k}$, $\vec{b}=\hat{i}+2\hat{j}+3\hat{k}$.

Solution We have $\vec{a}+\vec{b}=2\hat{i}+3\hat{j}+4\hat{k}$ and $\vec{a}-\vec{b}=-\hat{j}-2\hat{k}$

A vector which is perpendicular to both $\vec{a}+\vec{b}$ and $\vec{a}-\vec{b}$ is given by

$$(\vec{a}+\vec{b})\times(\vec{a}-\vec{b}) = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 2 & 3 & 4 \\ 0 & -1 & -2 \end{vmatrix} = -2\hat{i}+4\hat{j}-2\hat{k} \quad (=\vec{c}, \text{ say})$$

Now $$|\vec{c}| = \sqrt{4+16+4} = \sqrt{24} = 2\sqrt{6}$$

Therefore, the required unit vector is

$$\frac{\vec{c}}{|\vec{c}|} = \frac{-1}{\sqrt{6}}\hat{i}+\frac{2}{\sqrt{6}}\hat{j}-\frac{1}{\sqrt{6}}\hat{k}$$

Note There are two perpendicular directions to any plane. Thus, another unit vector perpendicular to $\vec{a}+\vec{b}$ and $\vec{a}-\vec{b}$ will be $\frac{1}{\sqrt{6}}\hat{i}-\frac{2}{\sqrt{6}}\hat{j}+\frac{1}{\sqrt{6}}\hat{k}$. But that will be a consequence of $(\vec{a}-\vec{b})\times(\vec{a}+\vec{b})$.

Example 24 Find the area of a triangle having the points A(1, 1, 1), B(1, 2, 3) and C(2, 3, 1) as its vertices.

Solution We have $\overrightarrow{AB}=\hat{j}+2\hat{k}$ and $\overrightarrow{AC}=\hat{i}+2\hat{j}$. The area of the given triangle is $\frac{1}{2}|\overrightarrow{AB}\times\overrightarrow{AC}|$.

Now, $$\overrightarrow{AB}\times\overrightarrow{AC} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{vmatrix} = -4\hat{i}+2\hat{j}-\hat{k}$$

Therefore $$|\overrightarrow{AB}\times\overrightarrow{AC}| = \sqrt{16+4+1} = \sqrt{21}$$

Thus, the required area is $\frac{1}{2}\sqrt{21}$

© NCERT not to be republished

Example 25 Find the area of a parallelogram whose adjacent sides are given by the vectors $\vec{a} = 3\hat{i} + \hat{j} + 4\hat{k}$ and $\vec{b} = \hat{i} - \hat{j} + \hat{k}$

Solution The area of a parallelogram with $\vec{a}$ and $\vec{b}$ as its adjacent sides is given by $|\vec{a} \times \vec{b}|$.

Now $$\vec{a} \times \vec{b} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ 3 & 1 & 4 \\ 1 & -1 & 1 \end{vmatrix} = 5\hat{i} + \hat{j} - 4\hat{k}$$

Therefore $$|\vec{a} \times \vec{b}| = \sqrt{25+1+16} = \sqrt{42}$$

and hence, the required area is $\sqrt{42}$.

EXERCISE 10.4

1. Find $|\vec{a} \times \vec{b}|$, if $\vec{a} = \hat{i} - 7\hat{j} + 7\hat{k}$ and $\vec{b} = 3\hat{i} - 2\hat{j} + 2\hat{k}$.
2. Find a unit vector perpendicular to each of the vector $\vec{a} + \vec{b}$ and $\vec{a} - \vec{b}$, where $\vec{a} = 3\hat{i} + 2\hat{j} + 2\hat{k}$ and $\vec{b} = \hat{i} + 2\hat{j} - 2\hat{k}$.
3. If a unit vector $\vec{a}$ makes angles $\frac{\pi}{3}$ with $\hat{i}$, $\frac{\pi}{4}$ with $\hat{j}$ and an acute angle θ with $\hat{k}$, then find θ and hence, the components of $\vec{a}$.
4. Show that
$$(\vec{a} - \vec{b}) \times (\vec{a} + \vec{b}) = 2(\vec{a} \times \vec{b})$$
5. Find λ and μ if $(2\hat{i} + 6\hat{j} + 27\hat{k}) \times (\hat{i} + \lambda\hat{j} + \mu\hat{k}) = \vec{0}$.
6. Given that $\vec{a} \neq \vec{0}$ and $\vec{a} \times \vec{b} = \vec{0}$. What can you conclude about the vectors $\vec{a}$ and $\vec{b}$?
7. Let the vectors $\vec{a}, \vec{b}, \vec{c}$ be given as $a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$, $b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$, $c_1\hat{i} + c_2\hat{j} + c_3\hat{k}$. Then show that $\vec{a} \times (\vec{b} + \vec{c}) = \vec{a} \times \vec{b} + \vec{a} \times \vec{c}$.
8. If either $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$, then $\vec{a} \times \vec{b} = \vec{0}$. Is the converse true? Justify your answer with an example.
9. Find the area of the triangle with vertices A(1, 1, 2), B(2, 3, 5) and C(1, 5, 5).

© NCERT not to be republished

10. Find the area of the parallelogram whose adjacent sides are determined by the vectors $\vec{a} = \hat{i} - \hat{j} + 3\hat{k}$ and $\vec{b} = 2\hat{i} - 7\hat{j} + \hat{k}$.

11. Let the vectors $\vec{a}$ and $\vec{b}$ be such that $|\vec{a}| = 3$ and $|\vec{b}| = \frac{\sqrt{2}}{3}$, then $\vec{a} \times \vec{b}$ is a unit vector, if the angle between $\vec{a}$ and $\vec{b}$ is

(A) $\pi/6$ (B) $\pi/4$ (C) $\pi/3$ (D) $\pi/2$

12. Area of a rectangle having vertices A, B, C and D with position vectors $-\hat{i} + \frac{1}{2}\hat{j} + 4\hat{k}$, $\hat{i} + \frac{1}{2}\hat{j} + 4\hat{k}$, $\hat{i} - \frac{1}{2}\hat{j} + 4\hat{k}$ and $-\hat{i} - \frac{1}{2}\hat{j} + 4\hat{k}$, respectively is

(A) $\frac{1}{2}$ (B) 1

(C) 2 (D) 4

Miscellaneous Examples

Example 26 Write all the unit vectors in XY-plane.

Solution Let $\vec{r} = x\hat{i} + y\hat{j}$ be a unit vector in XY-plane (Fig 10.28). Then, from the figure, we have $x = \cos\theta$ and $y = \sin\theta$ (since $|\vec{r}| = 1$). So, we may write the vector $\vec{r}$ as

$$\vec{r}\,(= \overrightarrow{OP}) = \cos\theta\,\hat{i} + \sin\theta\,\hat{j} \quad \ldots (1)$$

Clearly, $$|\vec{r}| = \sqrt{\cos^2\theta + \sin^2\theta} = 1$$

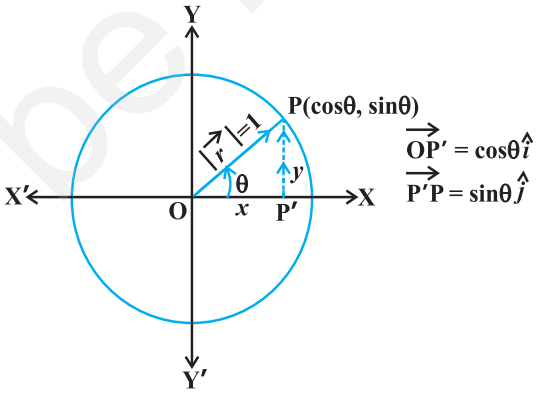


Fig 10.28

Also, as θ varies from 0 to 2π, the point P (Fig 10.28) traces the circle $x^2 + y^2 = 1$ counterclockwise, and this covers all possible directions. So, (1) gives every unit vector in the XY-plane.

© NCERT not to be republished

Example 27 If $\hat{i}\ \ \hat{j}\ \ \hat{k},\ 2\hat{i}\ \ 5\hat{j},\ 3\hat{i}\ \ 2\hat{j}\ \ 3\hat{k}$ and $\hat{i}\ \ 6\hat{j}\ \ \hat{k}$ are the position vectors of points A, B, C and D respectively, then find the angle between $\overrightarrow{AB}$ and $\overrightarrow{CD}$. Deduce that $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are collinear.

Solution Note that if θ is the angle between AB and CD, then θ is also the angle between $\overrightarrow{AB}$ and $\overrightarrow{CD}$.

Now $\overrightarrow{AB}$ = Position vector of B – Position vector of A

$$= (2\hat{i} + 5\hat{j}) - (\hat{i} + \hat{j} + \hat{k}) = \hat{i} + 4\hat{j} - \hat{k}$$

Therefore $|\overrightarrow{AB}| = \sqrt{(1)^2 + (4)^2 + (-1)^2} = 3\sqrt{2}$

Similarly $\overrightarrow{CD} = -2\hat{i} - 8\hat{j} + 2\hat{k}$ and $|\overrightarrow{CD}| = 6\sqrt{2}$

Thus $$\cos\theta = \frac{\overrightarrow{AB}\ \overrightarrow{CD}}{|\overrightarrow{AB}||\overrightarrow{CD}|}$$

$$= \frac{1(-2) + 4(-8) + (-1)(2)}{(3\sqrt{2})(6\sqrt{2})} = \frac{-36}{36} = -1$$

Since $0 \le \theta \le \pi$, it follows that $\theta = \pi$. This shows that $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are collinear.

Alternatively, $\overrightarrow{AB}\ \ \frac{1}{2}\overrightarrow{CD}$ which implies that $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are collinear vectors.

Example 28 Let $\vec{a}, \vec{b}$ and $\vec{c}$ be three vectors such that $|\vec{a}| = 3, |\vec{b}| = 4, |\vec{c}| = 5$ and each one of them being perpendicular to the sum of the other two, find $|\vec{a} + \vec{b} + \vec{c}|$.

Solution Given $\vec{a} \cdot (\vec{b} + \vec{c}) = 0,\ \vec{b} \cdot (\vec{c} + \vec{a}) = 0,\ \vec{c} \cdot (\vec{a} + \vec{b}) = 0.$

Now $|\vec{a} + \vec{b} + \vec{c}|^2 = (\vec{a} + \vec{b} + \vec{c})^2 = (\vec{a} + \vec{b} + \vec{c}) \cdot (\vec{a} + \vec{b} + \vec{c})$

$$= \vec{a} \cdot \vec{a} + \vec{a} \cdot (\vec{b} + \vec{c}) + \vec{b} \cdot \vec{b} + \vec{b} \cdot (\vec{a} + \vec{c})$$

$$+ \vec{c}.(\vec{a} + \vec{b}) + \vec{c}.\vec{c}$$

$$= |\vec{a}|^2 + |\vec{b}|^2 + |\vec{c}|^2$$

$$= 9 + 16 + 25 = 50$$

Therefore $|\vec{a} + \vec{b} + \vec{c}| = \sqrt{50} = 5\sqrt{2}$

© NCERT not to be republished

Example 29 Three vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ satisfy the condition $\vec{a}+\vec{b}+\vec{c}=\vec{0}$. Evaluate the quantity $\mu=\vec{a}\cdot\vec{b}+\vec{b}\cdot\vec{c}+\vec{c}\cdot\vec{a}$, if $|\vec{a}|=1$, $|\vec{b}|=4$ and $|\vec{c}|=2$.

Solution Since $\vec{a}+\vec{b}+\vec{c}=\vec{0}$, we have

$$\vec{a}\ (\vec{a}\quad\vec{b}\quad\vec{c})=0$$

or $$\vec{a}\cdot\vec{a}+\vec{a}\cdot\vec{b}+\vec{a}\cdot\vec{c}=0$$

Therefore $$\vec{a}\cdot\vec{b}+\vec{a}\cdot\vec{c}=-|\vec{a}|^2=-1 \qquad \text{... (1)}$$

Again, $$\vec{b}\cdot(\vec{a}+\vec{b}+\vec{c})=0$$

or $$\vec{a}\cdot\vec{b}+\vec{b}\cdot\vec{c}=-|\vec{b}|^2=-16 \qquad \text{... (2)}$$

Similarly $$\vec{a}\cdot\vec{c}+\vec{b}\cdot\vec{c}=-4. \qquad \text{... (3)}$$

Adding (1), (2) and (3), we have

$$2\,(\vec{a}\cdot\vec{b}+\vec{b}\cdot\vec{c}+\vec{a}\cdot\vec{c})=-21$$

or $$2\mu=-21, \text{ i.e., } \mu=\frac{-21}{2}$$

Example 30 If with reference to the right handed system of mutually perpendicular unit vectors $\hat{i}$, $\hat{j}$ and $\hat{k}$, $\vec{\alpha}=3\hat{i}-\hat{j}$, $\vec{\beta}=2\hat{i}+\hat{j}-3\hat{k}$, then express $\vec{\beta}$ in the form $\vec{\ }_1\ \vec{\ }_2$, where $\vec{\ }_1$ is parallel to $\vec{\ }$ and $\vec{\ }_2$ is perpendicular to $\vec{\alpha}$.

Solution Let $\vec{\ }_1\quad\vec{\ }$, is a scalar, i.e., $\vec{\beta}_1=3\lambda\hat{i}-\lambda\hat{j}$.

Now $$\vec{\beta}_2=\vec{\beta}-\vec{\beta}_1=(2-3\lambda)\hat{i}+(1+\lambda)\hat{j}-3\hat{k}.$$

Now, since $\vec{\beta}_2$ is to be perpendicular to $\vec{\alpha}$, we should have $\vec{\alpha}\cdot\vec{\beta}_2=0$. i.e.,

$$3(2-3\lambda)-(1+\lambda)=0$$

or $$\lambda=\frac{1}{2}$$

Therefore $$\vec{\beta}_1=\frac{3}{2}\hat{i}-\frac{1}{2}\hat{j} \quad \text{and} \quad \vec{\beta}_2=\frac{1}{2}\hat{i}+\frac{3}{2}\hat{j}-3\hat{k}$$

© NCERT not to be republished

Miscellaneous Exercise

1. Write down a unit vector in XY-plane, making an angle of 30° with the positive direction of x-axis.
2. Find the scalar components and magnitude of the vector joining the points $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$.
3. A girl walks 4 km towards west, then she walks 3 km in a direction 30° east of north and stops. Determine the girl's displacement from her initial point of departure.
4. If $\vec{a} = \vec{b} + \vec{c}$, then is it true that $|\vec{a}| = |\vec{b}| + |\vec{c}|$? Justify your answer.
5. Find the value of x for which $x(\hat{i} + \hat{j} + \hat{k})$ is a unit vector.
6. Find a vector of magnitude 5 units, and parallel to the resultant of the vectors $\vec{a} = 2\hat{i} + 3\hat{j} - \hat{k}$ and $\vec{b} = \hat{i} - 2\hat{j} + \hat{k}$.
7. If $\vec{a} = \hat{i} + \hat{j} + \hat{k}$, $\vec{b} = 2\hat{i} - \hat{j} + 3\hat{k}$ and $\vec{c} = \hat{i} - 2\hat{j} + \hat{k}$, find a unit vector parallel to the vector $2\vec{a} - \vec{b} + 3\vec{c}$.
8. Show that the points A(1, – 2, – 8), B(5, 0, – 2) and C(11, 3, 7) are collinear, and find the ratio in which B divides AC.
9. Find the position vector of a point R which divides the line joining two points P and Q whose position vectors are $(2\vec{a} + \vec{b})$ and $(\vec{a} - 3\vec{b})$ externally in the ratio 1 : 2. Also, show that P is the mid point of the line segment RQ.
10. The two adjacent sides of a parallelogram are $2\hat{i} - 4\hat{j} + 5\hat{k}$ and $\hat{i} - 2\hat{j} - 3\hat{k}$. Find the unit vector parallel to its diagonal. Also, find its area.
11. Show that the direction cosines of a vector equally inclined to the axes OX, OY and OZ are $\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}$.
12. Let $\vec{a} = \hat{i} + 4\hat{j} + 2\hat{k}$, $\vec{b} = 3\hat{i} - 2\hat{j} + 7\hat{k}$ and $\vec{c} = 2\hat{i} - \hat{j} + 4\hat{k}$. Find a vector $\vec{d}$ which is perpendicular to both $\vec{a}$ and $\vec{b}$, and $\vec{c} \cdot \vec{d} = 15$.
13. The scalar product of the vector $\hat{i} + \hat{j} + \hat{k}$ with a unit vector along the sum of vectors $2\hat{i} + 4\hat{j} - 5\hat{k}$ and $\lambda\hat{i} + 2\hat{j} + 3\hat{k}$ is equal to one. Find the value of λ.
14. If $\vec{a}$, $\vec{b}$, $\vec{c}$ are mutually perpendicular vectors of equal magnitudes, show that the vector $\vec{a} + \vec{b} + \vec{c}$ is equally inclined to $\vec{a}$, $\vec{b}$ and $\vec{c}$.

© NCERT not to be republished

15. Prove that $(\vec{a}+\vec{b})\cdot(\vec{a}+\vec{b}) = |\vec{a}|^2 + |\vec{b}|^2$, if and only if $\vec{a}, \vec{b}$ are perpendicular, given $\vec{a} \neq \vec{0}, \vec{b} \neq \vec{0}$.

Choose the correct answer in Exercises 16 to 19.

16. If θ is the angle between two vectors $\vec{a}$ and $\vec{b}$, then $\vec{a}\cdot\vec{b} \geq 0$ only when

(A) $0 < \theta < \frac{\pi}{2}$ (B) $0 \leq \theta \leq \frac{\pi}{2}$

(C) $0 < \theta < \pi$ (D) $0 \leq \theta \leq \pi$

17. Let $\vec{a}$ and $\vec{b}$ be two unit vectors and θ is the angle between them. Then $\vec{a}+\vec{b}$ is a unit vector if

(A) $\theta = \frac{\pi}{4}$ (B) $\theta = \frac{\pi}{3}$ (C) $\theta = \frac{\pi}{2}$ (D) $\theta = \frac{2\pi}{3}$

18. The value of $\hat{i}.(\hat{j} \times \hat{k}) + \hat{j}\cdot(\hat{i} \times \hat{k}) + \hat{k}\cdot(\hat{i} \times \hat{j})$ is

(A) 0 (B) –1 (C) 1 (D) 3

19. If θ is the angle between any two vectors $\vec{a}$ and $\vec{b}$, then $|\vec{a}\cdot\vec{b}| = |\vec{a}\times\vec{b}|$ when θ is equal to

(A) 0 (B) $\frac{\pi}{4}$ (C) $\frac{\pi}{2}$ (D) π

Summary

- Position vector of a point P(x, y, z) is given as $\overrightarrow{OP}(=\vec{r}) = x\hat{i} + y\hat{j} + z\hat{k}$, and its magnitude by $\sqrt{x^2+y^2+z^2}$.
- The scalar components of a vector are its direction ratios, and represent its projections along the respective axes.
- The magnitude (r), direction ratios (a, b, c) and direction cosines (l, m, n) of any vector are related as:

$$l = \frac{a}{r}, \quad m = \frac{b}{r}, \quad n = \frac{c}{r}$$

- The vector sum of the three sides of a triangle taken in order is $\vec{0}$.

© NCERT not to be republished

- The vector sum of two coinitial vectors is given by the diagonal of the parallelogram whose adjacent sides are the given vectors.
- The multiplication of a given vector by a scalar λ, changes the magnitude of the vector by the multiple $|\lambda|$, and keeps the direction same (or makes it opposite) according as the value of λ is positive (or negative).
- For a given vector $\vec{a}$, the vector $\hat{a} = \frac{\vec{a}}{|\vec{a}|}$ gives the unit vector in the direction of $\vec{a}$.
- The position vector of a point R dividing a line segment joining the points P and Q whose position vectors are $\vec{a}$ and $\vec{b}$ respectively, in the ratio $m : n$

  (i) internally, is given by $\frac{n\vec{a} + m\vec{b}}{m+n}$.

  (ii) externally, is given by $\frac{m\vec{b} - n\vec{a}}{m-n}$.
- The scalar product of two given vectors $\vec{a}$ and $\vec{b}$ having angle θ between them is defined as

$$\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos\theta .$$

  Also, when $\vec{a} \cdot \vec{b}$ is given, the angle 'θ' between the vectors $\vec{a}$ and $\vec{b}$ may be determined by

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|}$$

- If θ is the angle between two vectors $\vec{a}$ and $\vec{b}$, then their cross product is given as

$$\vec{a} \times \vec{b} = |\vec{a}||\vec{b}| \sin\theta \hat{n}$$

  where $\hat{n}$ is a unit vector perpendicular to the plane containing $\vec{a}$ and $\vec{b}$. Such that $\vec{a}, \vec{b}, \hat{n}$ form right handed system of coordinate axes.
- If we have two vectors $\vec{a}$ and $\vec{b}$, given in component form as $\vec{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$ and $\vec{b} = b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$ and λ any scalar,

© NCERT not to be republished

then $\vec{a}+\vec{b} = (a_1+b_1)\hat{i}+(a_2+b_2)\hat{j}+(a_3+b_3)\hat{k}$;

$\lambda\vec{a} = (\lambda a_1)\hat{i}+(\lambda a_2)\hat{j}+(\lambda a_3)\hat{k}$;

$\vec{a}.\vec{b} = a_1b_1+a_2b_2+a_3b_3$;

and $\vec{a}\times\vec{b} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix}$.

Historical Note

The word *vector* has been derived from a Latin word *vectus*, which means "to carry". The germinal ideas of modern vector theory date from around 1800 when Caspar Wessel (1745-1818) and Jean Robert Argand (1768-1822) described that how a complex number $a + ib$ could be given a geometric interpretation with the help of a directed line segment in a coordinate plane. William Rowen Hamilton (1805-1865) an Irish mathematician was the first to use the term vector for a directed line segment in his book *Lectures on Quaternions* (1853). Hamilton's method of quaternions (an ordered set of four real numbers given as: $a+b\hat{i}+c\hat{j}+d\hat{k}, \hat{i}, \hat{j}, \hat{k}$ following certain algebraic rules) was a solution to the problem of multiplying vectors in three dimensional space. Though, we must mention here that in practice, the idea of vector concept and their addition was known much earlier ever since the time of Aristotle (384-322 B.C.), a Greek philosopher, and pupil of Plato (427-348 B.C.). That time it was supposed to be known that the combined action of two or more forces could be seen by adding them according to parallelogram law. The correct law for the composition of forces, that forces add vectorially, had been discovered in the case of perpendicular forces by Stevin-Simon (1548-1620). In 1586 A.D., he analysed the principle of geometric addition of forces in his treatise *DeBeghinselen der Weeghconst* ("Principles of the Art of Weighing"), which caused a major breakthrough in the development of mechanics. But it took another 200 years for the general concept of vectors to form.

In the 1880, Josaih Willard Gibbs (1839-1903), an American physicist and mathematician, and Oliver Heaviside (1850-1925), an English engineer, created what we now know as *vector analysis*, essentially by separating the real (*scalar*)

© NCERT not to be republished

part of quaternion from its imaginary (*vector*) part. In 1881 and 1884, Gibbs printed a treatise entitled *Element of Vector Analysis*. This book gave a systematic and concise account of vectors. However, much of the credit for demonstrating the applications of vectors is due to the D. Heaviside and P.G. Tait (1831-1901) who contributed significantly to this subject.

Also this chapter contains 8 lectures:

Lecture 1

Vector Algebra

1.1 Vectors

Many physical quantities, such a mass, time, temperature are fully specified by one number or magnitude. They are **scalars**. But other quantities require more than one number to describe them. They are **vectors**. You have already met vectors in their more pure mathematical sense in your course on linear algebra (matrices and vectors), but often in the physical world, these numbers specify a **magnitude** and a **direction** — a total of two numbers in a 2D planar world, and three numbers in 3D.

Obvious examples are velocity, acceleration, electric field, and force. Below, probably all our examples will be of these "magnitude and direction" vectors, but we should not forget that many of the results extend to the wider realm of vectors.

There are three slightly different types of vectors:

- **Free vectors:** In many situtations *only* the magnitude and direction of a vector are important, and we can *translate* them at will (with 3 degrees of freedom for a vector in 3-dimensions).
- **Sliding vectors:** In mechanics the line of action of a force is often important for deriving moments. The force vector can slide with 1 degree of freedom.
- **Bound or position vectors:** When describing lines, curves etc in space, it is obviously important that the origin and head of the vector are not translated about arbitrarily. The origins of position vectors all coincide at an overall origin O.

One the advantages of using vectors is that it frees much of the analysis from the restriction of arbitrarily imposed coordinate frames. For example, if two free vectors are equal we need only say that their magnitudes and directions are equal, and that can be done with a drawing that is independent of any coordinate system.

However, coordinate systems are ultimately useful, so it useful to introduce the idea of vector components. Try to spot things in the notes that are independent

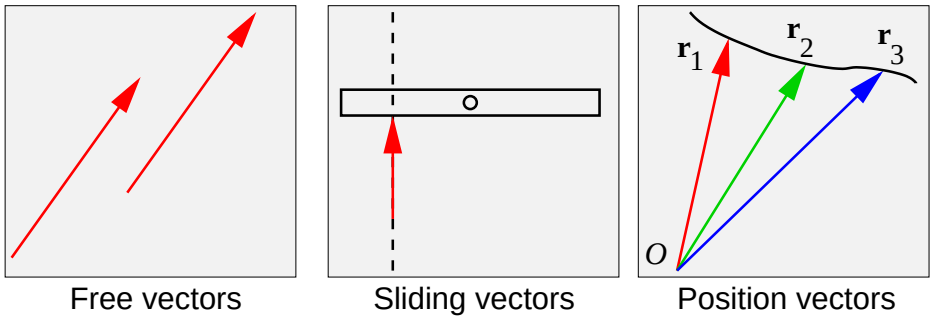


Figure 1.1:

of coordinate system.

1.1.1 Vector elements or components in a coordinate frame

A method of representing a vector is to list the values of its elements or components in a sufficient number of different (preferably mutually perpendicular) directions, depending on the dimension of the vector. These specified directions define a **coordinate frame**. In this course we will mostly restrict our attention to the 3-dimensional Cartesian coordinate frame $O(x, y, z)$. When we come to examine vector fields later in the course you will use curvilinear coordinate frames, especially 3D spherical and cylindrical polars, and 2D plane polar, coordinate systems.

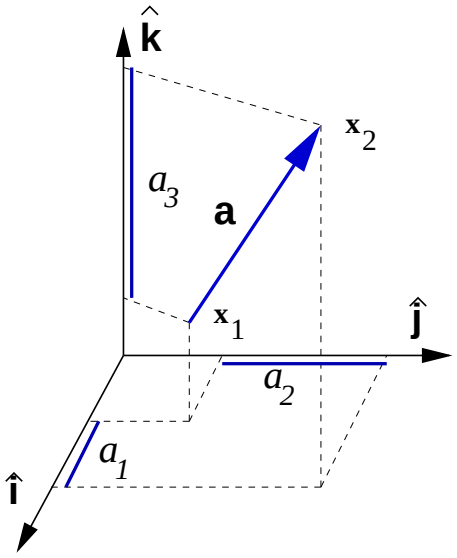


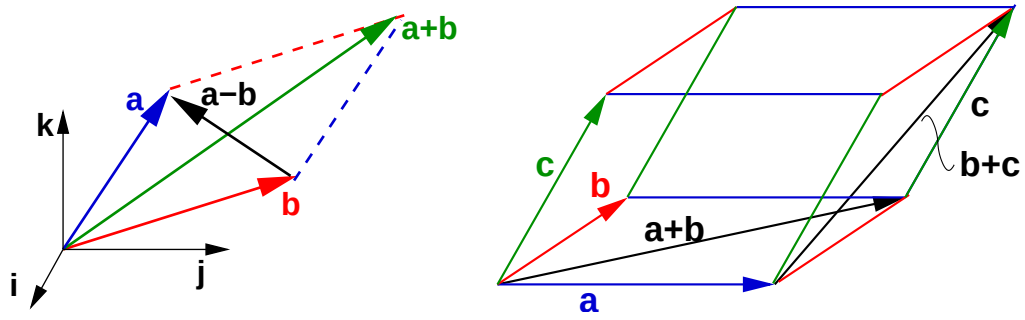
Figure 1.2: Vector components.

In a Cartesian coordinate frame we write

$$\mathbf{a} = [a_1, a_2, a_3] = [x_2 - x_1, y_2 - y_1, z_2 - z_1] \quad \text{or} \quad \mathbf{a} = [a_x, a_y, a_z]$$

as sketched in Figure 1.2. Defining $\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}$ as unit vectors in the x, y, z directions

$$\hat{\boldsymbol{\imath}} = [1, 0, 0] \quad \hat{\boldsymbol{\jmath}} = [0, 1, 0] \quad \hat{\boldsymbol{k}} = [0, 0, 1]$$



(a) (b)

Figure 1.3: (a) Addition of two vectors is commutative, but subtraction isn't. Note that the coordinate frame is irrelevant. (b) Addition of three vectors is associative.

we could also write

$$\mathbf{a} = a_1\hat{\boldsymbol{\imath}} + a_2\hat{\boldsymbol{\jmath}} + a_3\hat{\boldsymbol{k}} \quad .$$

Although we will be most often dealing with vectors in 3-space, you should not think that general vectors are limited to three components.

In these notes we will use bold font to represent vectors $\mathbf{a}, \boldsymbol{\omega}$, In your written work, underline the vector symbol $\underline{a}, \underline{\omega}$ and be **meticulous** about doing so. We shall use the hat to denote a unit vector.

1.1.2 Vector equality

Two free vectors are said to be equal iff their lengths and directions are the same. If we use a coordinate frame, we might say that corresponding components of the two vectors must be equal. This definition of equality will also do for position vectors, but for sliding vectors we must add that the line of action must be identical too.

1.1.3 Vector magnitude and unit vectors

Provided we use an orthogonal coordinate system, the magnitude of a 3-vector is

$$a = |\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

To find the unit vector in the direction of $\mathbf{a}$, simply divide by its magnitude

$$\hat{\mathbf{a}} = \frac{\mathbf{a}}{|\mathbf{a}|} \ .$$

1.1.4 Vector Addition and subtraction

Vectors are added/subtracted by adding/subtracting corresponding components, exactly as for matrices. Thus

$$\mathbf{a} + \mathbf{b} = [a_1 + b_1, \quad a_2 + b_2, \quad a_3 + b_3]$$

Addition follows the parallelogram construction of Figure 1.3(a). Subtraction ($\mathbf{a}-\mathbf{b}$) is defined as the addition ($\mathbf{a}+(-\mathbf{b})$). It is useful to remember that the vector $\mathbf{a}-\mathbf{b}$ goes from $\mathbf{b}$ *to* $\mathbf{a}$.

The following results follow immediately from the above definition of vector addition:

(a) $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ (commutativity) (Figure 1.3(a))

(b) $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})=\mathbf{a}+\mathbf{b}+\mathbf{c}$ (associativity) (Figure 1.3(b))

(c) $\mathbf{a}+\mathbf{0}=\mathbf{0}+\mathbf{a}=\mathbf{a}$, where the zero vector is $\mathbf{0}=[0,0,0]$.

(d) $\mathbf{a}+$(-$\mathbf{a}$)$=\mathbf{0}$

1.1.5 Multiplication of a vector by a scalar. (NOT the scalar product!)

Just as for matrices, multiplication of a vector $\mathbf{a}$ by a scalar c is defined as multiplication of each component by c, so that

$$c\mathbf{a}=[ca_1,ca_2,ca_3].$$

It follows that:

$$|c\mathbf{a}|=\sqrt{(ca_1)^2+(ca_2)^2+(ca_3)^2}=|c||\mathbf{a}|.$$

The direction of the vector will reverse if c is negative, but otherwise is unaffected. (By the way, a vector where the sign is uncertain is called a director.)

♣ Example

Q. Coulomb's law states that the electrostatic force on charged particle Q due to another charged particle q_1 is

$$\mathbf{F}=K\frac{Qq_1}{r^2}\hat{\mathbf{e}}_r$$

where $\mathbf{r}$ is the vector from q_1 to Q and $\hat{\mathbf{r}}$ is the unit vector in that same direction. (Note that the rule "unlike charges attract, like charges repel" is built into this formula.) The force between two particles is not modified by the presence of other charged particles.

Hence write down an expression for the force on Q at $\mathbf{R}$ due to N charges q_i at $\mathbf{r}_i$.

A. The vector from q_i to Q is $\mathbf{R}-\mathbf{r}_i$. The unit vector in that direction is $(\mathbf{R}-\mathbf{r}_i)/|\mathbf{R}-\mathbf{r}_i|$, so the resultant force is

$$\mathbf{F}(\mathbf{R})=\sum_{i=1}^{N}K\frac{Qq_i}{|\mathbf{R}-\mathbf{r}_i|^3}(\mathbf{R}-\mathbf{r}_i)\quad .$$

Note that $\mathbf{F}(\mathbf{R})$ is a vector *field*.

1.2 Scalar, dot, or inner product

This is a product of two vectors results in a scalar quantity and is defined as follows for 3-component vectors:

$$\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3 \ .$$

Note that

$$\mathbf{a}\cdot\mathbf{a} = a_1^2 + a_2^2 + a_3^2 = |\mathbf{a}|^2 = a^2.$$

The following laws of multiplication follow immediately from the definition:

(a) $\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$ (commutativity)

(b) $\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}$ (distributivity with respect to vector addition)

(c) $(\lambda\mathbf{a})\cdot\mathbf{b} = \lambda(\mathbf{a}\cdot\mathbf{b}) = \mathbf{a}\cdot(\lambda\mathbf{b})$ scalar multiple of a scalar product of two vectors

1.2.1 Geometrical interpretation of scalar product

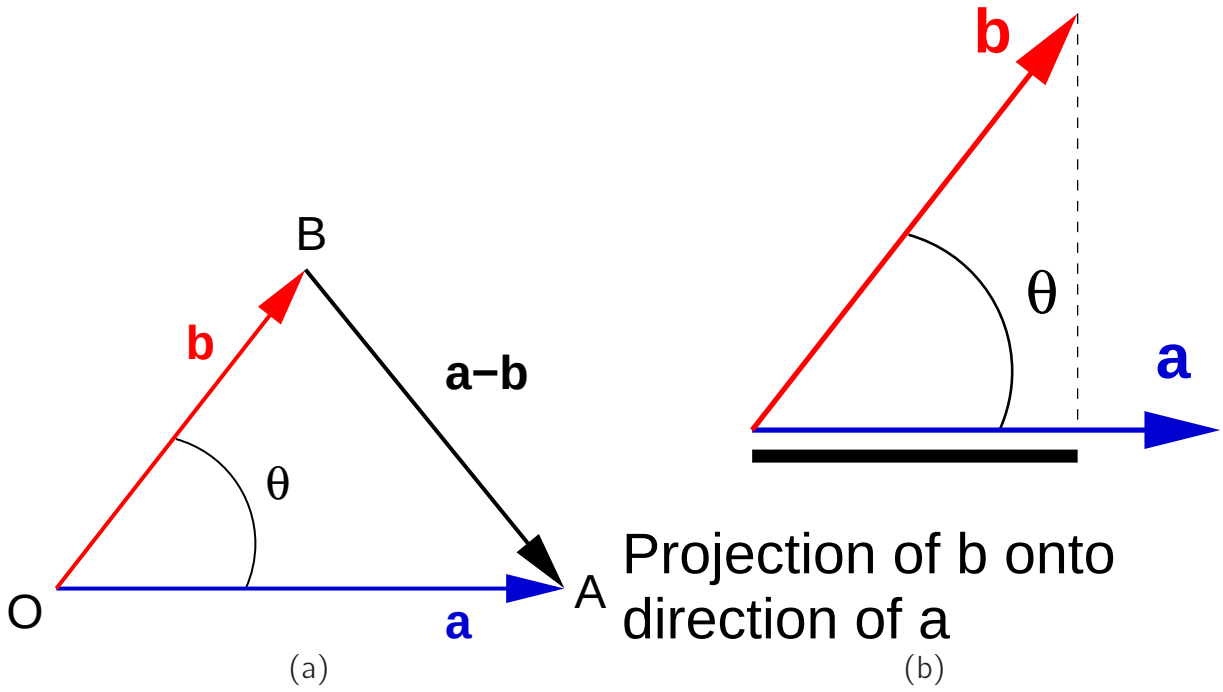


Figure 1.4: (a) Cosine rule. (b) Projection of **b** onto **a**.

Consider the square magnitude of the vector $\mathbf{a}-\mathbf{b}$. By the rules of the scalar product, this is

$$\begin{aligned} |\mathbf{a}-\mathbf{b}|^2 &= (\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b}) \\ &= \mathbf{a}\cdot\mathbf{a} + \mathbf{b}\cdot\mathbf{b} - 2(\mathbf{a}\cdot\mathbf{b}) \\ &= a^2 + b^2 - 2(\mathbf{a}\cdot\mathbf{b}) \end{aligned}$$

But, by the cosine rule for the triangle OAB (Figure 1.4a), the length AB^2 is given by

$$|\mathbf{a} - \mathbf{b}|^2 = a^2 + b^2 - 2ab\cos\theta$$

where θ is the angle between the two vectors. It follows that

$$\mathbf{a} \cdot \mathbf{b} = ab\cos\theta,$$

which is independent of the co-ordinate system used, and that $|\mathbf{a} \cdot \mathbf{b}| \leq ab$. Conversely, the cosine of the angle between vectors $\mathbf{a}$ and $\mathbf{b}$ is given by $\cos\theta = \mathbf{a}\cdot\mathbf{b}/ab$.

1.2.2 Projection of one vector onto the other

Another way of describing the scalar product is as the product of the magnitude of one vector and the component of the other in the direction of the first, since $b\cos\theta$ is the component of $\mathbf{b}$ in the direction of $\mathbf{a}$ and vice versa (Figure 1.4b).

Projection is particularly useful when the second vector is a unit vector — $\mathbf{a} \cdot \hat{\boldsymbol{\imath}}$ is the component of $\mathbf{a}$ in the direction of $\hat{\boldsymbol{\imath}}$.

Notice that if we wanted the **vector component** of $\mathbf{b}$ in the direction of $\mathbf{a}$ we would write

$$(\mathbf{b} \cdot \hat{\mathbf{a}})\hat{\mathbf{a}} = \frac{(\mathbf{b} \cdot \mathbf{a})\mathbf{a}}{a^2} \quad .$$

In the particular case $\mathbf{a} \cdot \mathbf{b} = 0$, the angle between the two vectors is a right angle and the vectors are said to be mutually orthogonal or perpendicular — neither vector has any component in the direction of the other.

An orthonormal coordinate system is characterised by $\hat{\boldsymbol{\imath}} \cdot \hat{\boldsymbol{\imath}} = \hat{\boldsymbol{\jmath}} \cdot \hat{\boldsymbol{\jmath}} = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}} = 1$; and $\hat{\boldsymbol{\imath}} \cdot \hat{\boldsymbol{\jmath}} = \hat{\boldsymbol{\jmath}} \cdot \hat{\boldsymbol{k}} = \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{\imath}} = 0$.

1.2.3 A scalar product is an "inner product"

So far we have been writing our vectors as row vectors $\mathbf{a} = [a_1,\ a_2,\ a_3]$. This is convenient because it takes up less room than writing column vectors

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad .$$

In matrix algebra vectors are more usually defined as column vectors, as in

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

and a row vector is written as $\mathbf{a}^\top$. Now for most of our work we can be quite relaxed about this minor difference, but here let us be fussy.

Why? Simply to point out at that the scalar product is also the **inner product** more commonly used in linear algebra. Defined as $\mathbf{a}^\top\mathbf{b}$ when vectors are column vectors as

$$\mathbf{a}\cdot\mathbf{b} = \mathbf{a}^\top\mathbf{b} = [a_1,\ a_2,\ a_3]\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1b_1 + a_2b_2 + a_3b_3\ .$$

Here we treat a n-dimensional column vector as an $n \times 1$ matrix.

(Remember that if you multiply two matrices $\mathsf{M}_{m\times n}\mathsf{N}_{n\times p}$ then M must have the same columns as N has rows (here denoted by n) and the result has size (rows × columns) of $m \times p$. So for n-dimensional column vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a}^\top$ is a $1 \times n$ matrix and $\mathbf{b}$ is $n \times 1$ matrix, so the product $\mathbf{a}^\top\mathbf{b}$ is a 1×1 matrix, which is (at last!) a scalar.)

♣ Examples

Q1. A force $\mathbf{F}$ is applied to an object as it moves by a small amount $\delta\mathbf{r}$. What work is done on the object by the force?

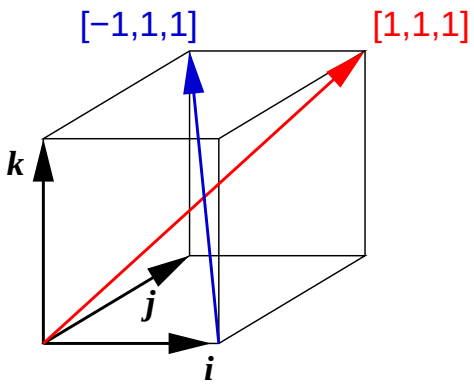
A1. The work done is equal to the component of force in the direction of the displacement multiplied by the displacement itself. This is just a scalar product:

$$\delta W = \mathbf{F}\cdot\delta\mathbf{r}\quad .$$

Q2. A cube has four diagonals, connecting opposite vertices. What is the angle between an adjacent pair?

A2. Well, you could plod through using Pythagoras' theorem to find the length of the diagonal from cube vertex to cube centre, and perhaps you should to check the following answer.

The directions of the diagonals are $[\pm1, \pm1, \pm1]$. The ones shown in the figure are $[1, 1, 1]$ and $[-1, 1, 1]$. The angle is thus

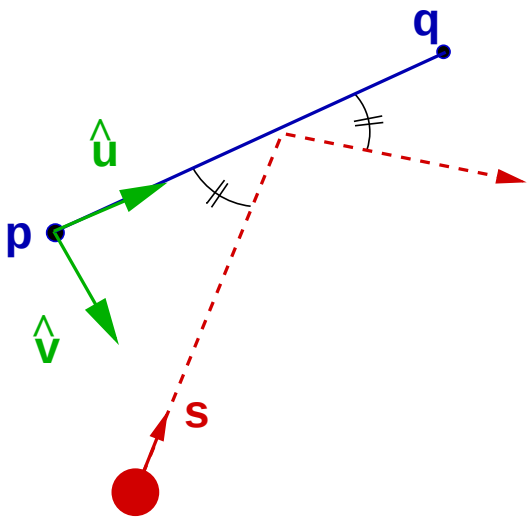


$$\theta = \cos^{-1}\frac{[1,1,1]\cdot[-1,1,1]}{\sqrt{1^2+1^2+1^2}\sqrt{-1^2+1^2+1^2}} = \cos^{-1}\frac{1}{3}$$

Q3. A pinball moving in a plane with velocity **s** bounces (in a purely elastic impact) from a baffle whose endpoints are **p** and **q**. What is the velocity vector after the bounce?

A3. Best to refer everything to a coordinate frame with principal directions $\hat{\mathbf{u}}$ along and $\hat{\mathbf{v}}$ perpendicular to the baffle:

$$\begin{aligned}\hat{\mathbf{u}} &= \frac{\mathbf{q}-\mathbf{p}}{|\mathbf{q}-\mathbf{p}|}\\ \hat{\mathbf{v}} &= \mathbf{u}^{\perp} = [-u_y, \quad u_x]\end{aligned}$$



Thus the velocity before impact is

$$\mathbf{s}_{\text{before}} = (\mathbf{s}.\hat{\mathbf{u}})\hat{\mathbf{u}} + (\mathbf{s}.\hat{\mathbf{v}})\hat{\mathbf{v}}$$

After the impact, the component of velocity in the direction of the baffle is unchanged and the component normal to the baffle is negated:

$$\mathbf{s}_{\text{after}} = (\mathbf{s}.\hat{\mathbf{u}})\hat{\mathbf{u}} - (\mathbf{s}.\hat{\mathbf{v}})\hat{\mathbf{v}}$$

1.2.4 Direction cosines use projection

Direction cosines are commonly used in the field of crystallography. The quantities

$$\lambda = \frac{\mathbf{a}\cdot\hat{\imath}}{a}, \qquad \mu = \frac{\mathbf{a}\cdot\hat{\jmath}}{a}, \qquad \nu = \frac{\mathbf{a}\cdot\hat{k}}{a}$$

represent the cosines of the angles which the vector **a** makes with the co-ordinate vectors $\hat{\imath}$,$\hat{\jmath}$, $\hat{k}$ and are known as the direction cosines of the vector **a**. Since $\mathbf{a}\cdot\hat{\imath} = a_1$ etc, it follows immediately that $\mathbf{a} = a(\lambda\hat{\imath} + \mu\hat{\jmath} + \nu\hat{k})$ and $\lambda^2 + \mu^2 + \nu^2 = \frac{1}{a^2}[a_1^2 + a_2^2 + a_3^2] = 1$

1.3 Vector or cross product

The vector product of two vectors **a** and **b** is denoted by $\mathbf{a}\times\mathbf{b}$ and is defined as follows

$$\mathbf{a}\times\mathbf{b} = (a_2b_3 - a_3b_2)\hat{\imath} + (a_3b_1 - a_1b_3)\hat{\jmath} + (a_1b_2 - a_2b_1)\hat{k}.$$

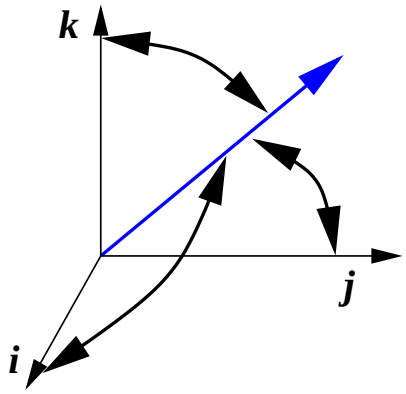


Figure 1.5: The direction cosines are cosines of the angles shown.

It is MUCH more easily remembered in terms of the (pseudo-)determinant

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{vmatrix}$$

where the top row consists of the vectors $\hat{\boldsymbol{\imath}}$, $\hat{\boldsymbol{\jmath}}$, $\hat{\boldsymbol{k}}$ rather than scalars.

Since a determinant with two equal rows has value zero, it follows that $\mathbf{a}\times\mathbf{a}=\mathbf{0}$. It is also easily verified that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{a}=(\mathbf{a}\times\mathbf{b})\cdot\mathbf{b}=0$, so that $\mathbf{a}\times\mathbf{b}$ is orthogonal (perpendicular) to both $\mathbf{a}$ and $\mathbf{b}$, as shown in Figure 1.6.

Note that $\hat{\boldsymbol{\imath}}\times\hat{\boldsymbol{\jmath}}=\hat{\boldsymbol{k}}$, $\hat{\boldsymbol{\jmath}}\times\hat{\boldsymbol{k}}=\hat{\boldsymbol{\imath}}$, and $\hat{\boldsymbol{k}}\times\hat{\boldsymbol{\imath}}=\hat{\boldsymbol{\jmath}}$.

The magnitude of the vector product can be obtained by showing that

$$|\mathbf{a}\times\mathbf{b}|^2+(\mathbf{a}\cdot\mathbf{b})^2=a^2b^2$$

from which it follows that

$$|\mathbf{a}\times\mathbf{b}|=ab\sin\theta\ ,$$

which is again independent of the co-ordinate system used. This is left as an exercise.

Unlike the scalar product, the vector product does not satisfy commutativity but is in fact anti-commutative, in that $\mathbf{a}\times\mathbf{b}=-\mathbf{b}\times\mathbf{a}$. Moreover the vector product does not satisfy the associative law of multiplication either since, as we shall see later $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\neq(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$.

Since the vector product is known to be orthogonal to both the vectors which form the product, it merely remains to specify its sense with respect to these vectors. Assuming that the co-ordinate vectors form a right-handed set in the order $\hat{\boldsymbol{\imath}}$,$\hat{\boldsymbol{\jmath}}$, $\hat{\boldsymbol{k}}$ it can be seen that the sense of the the vector product is also right handed, i.e

the vector product has the same sense as the co-ordinate system used.

$$\hat{\boldsymbol{\imath}} \times \hat{\boldsymbol{\jmath}} = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \hat{\boldsymbol{k}} \quad .$$

In practice, figure out the direction from a right-handed screw twisted from the first to second vector as shown in Figure 1.6(a).

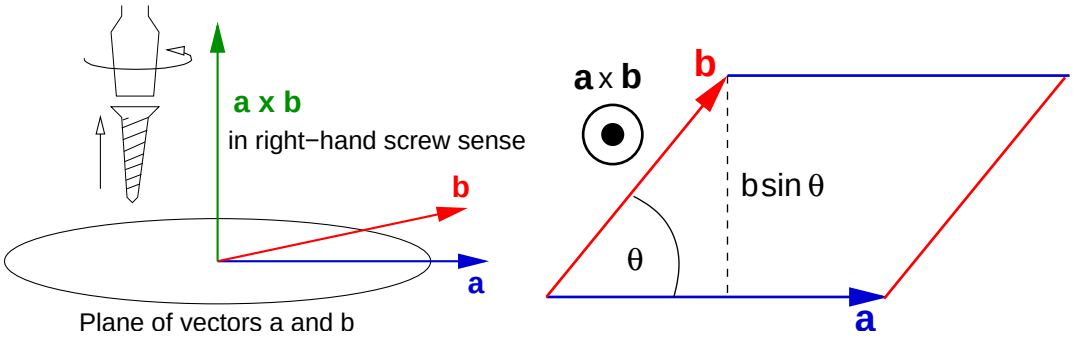


Figure 1.6: (a)The vector product is orthogonal to both **a** and **b**. Twist from first to second and move in the direction of a right-handed screw. (b) Area of parallelogram is $ab\sin\theta$.

1.3.1 Geometrical interpretation of vector product

The magnitude of the vector product $(\mathbf{a} \times \mathbf{b})$ is equal to the area of the parallelogram whose sides are parallel to, and have lengths equal to the magnitudes of, the vectors **a** and **b** (Figure 1.6b). Its direction is perpendicular to the parallelogram.

♣ Example

Q. **g** is vector from A [1,2,3] to B [3,4,5].
$\hat{\boldsymbol{\ell}}$ is the unit vector in the direction from O to A.
Find $\hat{\mathbf{m}}$, a UNIT vector along $\mathbf{g} \times \hat{\boldsymbol{\ell}}$
Verify that $\hat{\mathbf{m}}$ is perpendicular to $\hat{\boldsymbol{\ell}}$.
Find $\hat{\mathbf{n}}$, the third member of a right-handed coordinate set $\hat{\boldsymbol{\ell}}, \hat{\mathbf{m}}, \hat{\mathbf{n}}$.

A.

$$\mathbf{g} = [3,4,5] - [1,2,3] = [2,2,2]$$

$$\hat{\boldsymbol{\ell}} = \frac{1}{\sqrt{14}}[1,2,3]$$

$$\mathbf{g} \times \hat{\boldsymbol{\ell}} = \frac{1}{\sqrt{14}} \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 2 & 2 & 2 \\ 1 & 2 & 3 \end{vmatrix} = \frac{1}{\sqrt{14}}[2,-4,2]$$

Hence

$$\hat{\mathbf{m}} = \cdot\frac{1}{\sqrt{24}}[2, -4, 2]$$

and

$$\hat{\mathbf{n}} = \hat{\boldsymbol{\ell}} \times \hat{\mathbf{m}}$$

Lecture 2

Multiple Products. Geometry using Vectors

2.1 Triple and multiple products

Using mixtures of the pairwise scalar product and vector product, it is possible to derive "triple products" between three vectors, and indeed n-products between n vectors.

There is nothing about these that you cannot work out from the definitions of pairwise scalar and vector products already given, but some have interesting geometric interpretations, so it is worth looking at these.

2.1.1 Scalar triple product

This is the scalar product of a vector product and a third vector, i.e. $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. Using the pseudo-determinant expression for the vector product, we see that the scalar triple product can be represented as the true determinant

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

You will recall that if you swap a pair of rows of a determinant, its sign changes; hence if you swap two pairs, its sign stays the same.

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \quad \begin{matrix} \\ \\ + \end{matrix} \qquad \text{1st SWAP} \qquad \begin{vmatrix} c_1 & c_2 & c_3 \\ b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \end{vmatrix} \quad \begin{matrix} \\ \\ - \end{matrix} \qquad \text{2nd SWAP} \qquad \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \quad \begin{matrix} \\ \\ + \end{matrix}$$

This says that

(i) $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$ (Called cyclic permutation.)

(ii) $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = -\mathbf{b} \cdot (\mathbf{a} \times \mathbf{c})$ and so on. (Called anti-cyclic permutation.)

(iii) The fact that $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ allows the scalar triple product to be written as $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. This notation is not very helpful, and we will try to avoid it below.

2.1.2 Geometrical interpretation of scalar triple product

The scalar triple product gives the volume of the parallelopiped whose sides are represented by the vectors **a**, **b**, and **c**.

We saw earlier that the vector product $(\mathbf{a} \times \mathbf{b})$ has magnitude equal to the area of the base, and direction perpendicular to the base. The *component* of **c** in this direction is equal to the height of the parallelopiped shown in Figure 2.1(a).

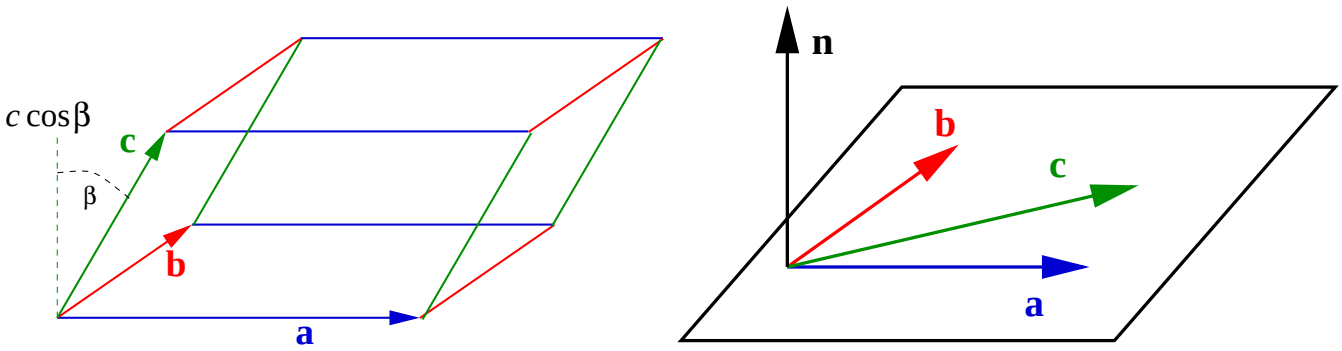


Figure 2.1: (a) Scalar triple product equals volume of parallelopiped. (b) Coplanarity yields zero scalar triple product.

2.1.3 Linearly dependent vectors

If the scalar triple product of three vectors is zero

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0$$

then the vectors are **linearly dependent**. That is, one can be expressed as a linear combination of the others. For example,

$$\mathbf{a} = \lambda \mathbf{b} + \mu \mathbf{c}$$

where λ and μ are scalar coefficients.

You can see this immediately in two ways:

- The determinant would have one row that was a linear combination of the others. You'll remember that by doing row operations, you could reach a row of zeros, and so the determinant is zero.
- The parallelopiped would have zero volume if squashed flat. In this case all three vectors lie in a plane, and so any one is a linear combination of the other two. (Figure 2.1b.)

2.1.4 Vector triple product

This is defined as the vector product of a vector with a vector product, $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$. Now, the vector triple product $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ must be perpendicular to $(\mathbf{b}\times\mathbf{c})$, which in turn is perpendicular to both $\mathbf{b}$ and $\mathbf{c}$. Thus $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ can have no component perpendicular to $\mathbf{b}$ and $\mathbf{c}$, and hence must be coplanar with them. It follows that the vector triple product must be expressible as a linear combination of $\mathbf{b}$ and $\mathbf{c}$:

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \lambda\mathbf{b} + \mu\mathbf{c}\ .$$

The values of the coefficients can be obtained by multiplying out in component form. Only the first component need be evaluated, the others then being obtained by symmetry. That is

$$\begin{aligned}(\mathbf{a}\times(\mathbf{b}\times\mathbf{c}))_1 &= a_2(\mathbf{b}\times\mathbf{c})_3 - a_3(\mathbf{b}\times\mathbf{c})_2\\ &= a_2(b_1c_2 - b_2c_1) + a_3(b_1c_3 - b_3c_1)\\ &= (a_2c_2 + a_3c_3)b_1 - (a_2b_2 + a_3b_3)c_1\\ &= (a_1c_1 + a_2c_2 + a_3c_3)b_1 - (a_1b_1 + a_2b_2 + a_3b_3)c_1\\ &= (\mathbf{a}\cdot\mathbf{c})b_1 - (\mathbf{a}\cdot\mathbf{b})c_1\end{aligned}$$

The equivalents must be true for the 2nd and 3rd components, so we arrive at the identity

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}\ .$$

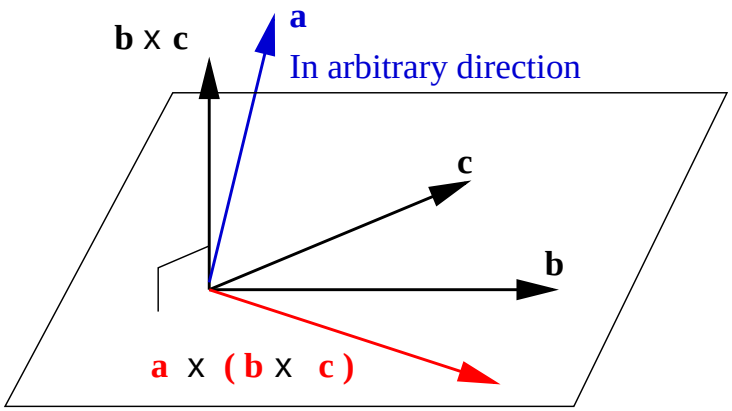


Figure 2.2: Vector triple product.

2.1.5 Projection using vector triple product

An example of the application of this formula is as follows. Suppose $\mathbf{v}$ is a vector and we want its projection into the xy-plane. The component of $\mathbf{v}$ in the z direction is $\mathbf{v}\cdot\hat{\boldsymbol{k}}$, so the projection we seek is $\mathbf{v} - (\mathbf{v}\cdot\hat{\boldsymbol{k}})\hat{\boldsymbol{k}}$. Writing $\hat{\boldsymbol{k}} \leftarrow \mathbf{a}$, $\mathbf{v} \leftarrow \mathbf{b}$,

$\hat{\boldsymbol{k}} \leftarrow \mathbf{c}$,

$$\begin{array}{rcl} \mathbf{a}\times(\mathbf{b}\times\mathbf{c}) &=& (\mathbf{a}\cdot\mathbf{c})\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\mathbf{c} \\ \downarrow \quad \downarrow \quad \downarrow & & \\ \hat{\boldsymbol{k}}\times(\mathbf{v}\times\hat{\boldsymbol{k}}) &=& (\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}})\mathbf{v}-(\hat{\boldsymbol{k}}\cdot\mathbf{v})\hat{\boldsymbol{k}} \\ &=& \mathbf{v}-(\mathbf{v}\cdot\hat{\boldsymbol{k}})\hat{\boldsymbol{k}} \end{array}$$

So $\mathbf{v}-(\mathbf{v}\cdot\hat{\boldsymbol{k}})\hat{\boldsymbol{k}} = \hat{\boldsymbol{k}}\times(\mathbf{v}\times\hat{\boldsymbol{k}})$.

(Hot stuff! But the expression $\mathbf{v}-(\mathbf{v}\cdot\hat{\boldsymbol{k}})\hat{\boldsymbol{k}}$ is much easier to understand, and cheaper to compute!)

2.1.6 Other repeated products

Many combinations of vector and scalar products are possible, but we consider only one more, namely the vector quadruple product $(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})$. By regarding $\mathbf{a}\times\mathbf{b}$ as a single vector, we see that this vector must be representable as a linear combination of $\mathbf{c}$ and $\mathbf{d}$. On the other hand, regarding $\mathbf{c}\times\mathbf{d}$ as a single vector, we see that it must also be a linear combination of $\mathbf{a}$ and $\mathbf{b}$. This provides a means of expressing one of the vectors, say $\mathbf{d}$, as linear combination of the other three, as follows:

$$\begin{array}{rcl} (\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d}) &=& [(\mathbf{a}\times\mathbf{b})\cdot\mathbf{d}]\mathbf{c}-[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]\mathbf{d} \\ &=& [(\mathbf{c}\times\mathbf{d})\cdot\mathbf{a}]\mathbf{b}-[(\mathbf{c}\times\mathbf{d})\cdot\mathbf{b}]\mathbf{a} \end{array}$$

Hence

$$[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]\,\mathbf{d} = [(\mathbf{b}\times\mathbf{c})\cdot\mathbf{d}]\,\mathbf{a}+[(\mathbf{c}\times\mathbf{a})\cdot\mathbf{d}]\,\mathbf{b}+[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{d}]\,\mathbf{c}$$

or

$$\mathbf{d} = \frac{[(\mathbf{b}\times\mathbf{c})\cdot\mathbf{d}]\,\mathbf{a}+[(\mathbf{c}\times\mathbf{a})\cdot\mathbf{d}]\,\mathbf{b}+[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{d}]\,\mathbf{c}}{[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]} \;=\; \alpha\mathbf{a}+\beta\mathbf{b}+\gamma\mathbf{c} \quad .$$

This is not something to remember off by heart, but it is worth remembering that the projection of a vector on any arbitrary basis set is unique.

♣ Example

Q1 Use the quadruple vector product to express the vector $\mathbf{d}=[3,2,1]$ in terms of the vectors $\mathbf{a}=[1,2,3]$, $\mathbf{b}=[2,3,1]$ and $\mathbf{c}=[3,1,2]$.

A1 Grinding away at the determinants, we find

$$[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]=-18;\;\; [(\mathbf{b}\times\mathbf{c})\cdot\mathbf{d}]=6;\;\; [(\mathbf{c}\times\mathbf{a})\cdot\mathbf{d}]=-12;\;\; [(\mathbf{a}\times\mathbf{b})\cdot\mathbf{d}]=-12$$

So, $\mathbf{d}=(-\mathbf{a}+2\mathbf{b}+2\mathbf{c})/3$.

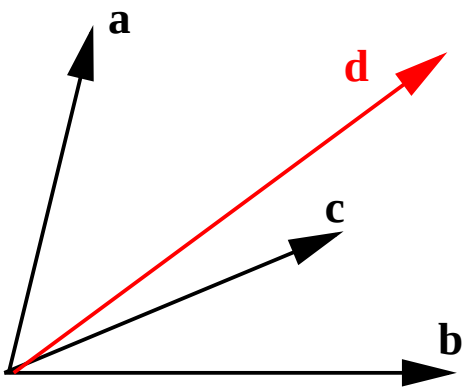


Figure 2.3: The projection of a (3-)vector onto a set of (3) basis vectors is unique. Ie in $\mathbf{d} = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$, the set $\{\alpha, \beta, \gamma\}$ is unique.

2.2 Geometry using vectors: lines, planes

2.2.1 The equation of a line

The equation of the line passing through the point whose position vector is $\mathbf{a}$ and lying in the direction of vector $\mathbf{b}$ is

$$\mathbf{r} = \mathbf{a} + \lambda\mathbf{b}$$

where λ is a scalar parameter. If you make $\mathbf{b}$ a unit vector, $\mathbf{r} = \mathbf{a} + \lambda\hat{\mathbf{b}}$ then λ will represent metric length.

For a line defined by two points $\mathbf{a}_1$ and $\mathbf{a}_2$

$$\mathbf{r} = \mathbf{a}_1 + \lambda(\mathbf{a}_2 - \mathbf{a}_1)$$

or for the unit version

$$\mathbf{r} = \mathbf{a}_1 + \lambda\frac{(\mathbf{a}_2 - \mathbf{a}_1)}{|\mathbf{a}_2 - \mathbf{a}_1|}$$

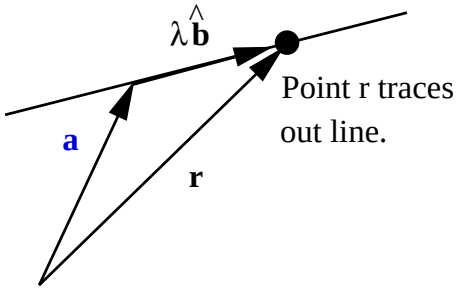


Figure 2.4: Equation of a line. With $\hat{\mathbf{b}}$ a unit vector, λ is in the length units established by the definition of $\mathbf{a}$.

2.2.2 The shortest distance from a point to a line

Referring to Figure 2.5(a) the vector $\mathbf{p}$ from $\mathbf{c}$ to any point on the line is $\mathbf{p} = \mathbf{a} + \lambda\hat{\mathbf{b}} - \mathbf{c} = (\mathbf{a} - \mathbf{c}) + \lambda\hat{\mathbf{b}}$ which has length squared $p^2 = (\mathbf{a} - \mathbf{c})^2 + \lambda^2 + 2\lambda(\mathbf{a} - \mathbf{c}) \cdot \hat{\mathbf{b}}$. Rather than minimizing length, it is easier to minimize length-squared. The minumum is found when $d\ p^2/d\lambda = 0$, ie when

$$\lambda = -(\mathbf{a} - \mathbf{c}) \cdot \hat{\mathbf{b}} \ .$$

So the minimum length vector is

$$\mathbf{p} = (\mathbf{a} - \mathbf{c}) - ((\mathbf{a} - \mathbf{c}) \cdot \hat{\mathbf{b}})\hat{\mathbf{b}}.$$

You might spot that is the component of $(\mathbf{a} - \mathbf{c})$ perpendicular to $\hat{\mathbf{b}}$ (as expected!). Furthermore, using the result of Section 2.1.5,

$$\mathbf{p} = \hat{\mathbf{b}} \times [(\mathbf{a} - \mathbf{c}) \times \hat{\mathbf{b}}] \ .$$

Because $\hat{\mathbf{b}}$ is a unit vector, and is orthogonal to $[(\mathbf{a} - \mathbf{c}) \times \hat{\mathbf{b}}]$, the modulus of the vector can be written rather more simply as just

$$p_{\text{min}} = |(\mathbf{a} - \mathbf{c}) \times \hat{\mathbf{b}}| \ .$$

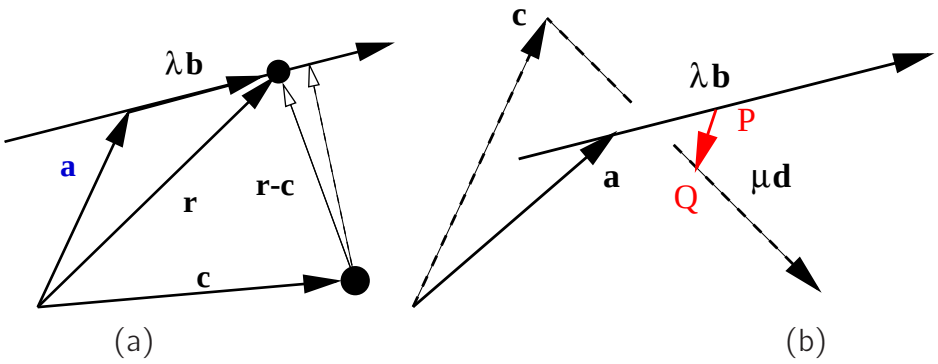


Figure 2.5: (a) Shortest distance point to line. (b) Shortest distance, line to line.

2.2.3 The shortest distance between two straight lines

If the shortest distance between a point and a line is along the perpendicular, then the shortest distance between the two straight lines $\mathbf{r} = \mathbf{a} + \lambda\hat{\mathbf{b}}$ and $\mathbf{r} = \mathbf{c} + \mu\hat{\mathbf{d}}$ must be found as the length of the vector which is mutually perpendicular to the lines.

The unit vector along the mutual perpendicular is

$$\hat{\mathbf{p}} = (\hat{\mathbf{b}} \times \hat{\mathbf{d}})/|\hat{\mathbf{b}} \times \hat{\mathbf{d}}| \ .$$

(Yes, don't forget that $\hat{\mathbf{b}} \times \hat{\mathbf{d}}$ is NOT a unit vector. $\hat{\mathbf{b}}$ and $\hat{\mathbf{d}}$ are not orthogonal, so there is a $\sin\theta$ lurking!)

The minimum length is therefore the component of $\mathbf{a} - \mathbf{c}$ in this direction

$$p_{\text{min}} = \left|(\mathbf{a} - \mathbf{c}) \cdot (\hat{\mathbf{b}} \times \hat{\mathbf{d}})/|\hat{\mathbf{b}} \times \hat{\mathbf{d}}|\right| \ .$$

♣ Example

Q Two long straight pipes are specified using Cartesian co-ordinates as follows: Pipe A has diameter 0.8 and its axis passes through points $(2, 5, 3)$ and $(7, 10, 8)$.
Pipe B has diameter 1.0 and its axis passes through the points $(0, 6, 3)$ and $(-12, 0, 9)$.
Determine whether the pipes need to be realigned to avoid intersection.

A Each pipe axis is defined using two points. The vector equation of the axis of pipe A is

$$\mathbf{r} = [2, 5, 3] + \lambda'[5, 5, 5] = [2, 5, 3] + \lambda[1, 1, 1]/\sqrt{3}$$

The equation of the axis of pipe B is

$$\mathbf{r} = [0, 6, 3] + \mu'[12, 6, 6] = [0, 6, 3] + \mu[-2, -1, 1]/\sqrt{6}$$

The perpendicular to the two axes has direction

$$[1, 1, 1] \times [-2, -1, 1] = \begin{vmatrix} \hat{\boldsymbol{i}} & \hat{\boldsymbol{j}} & \hat{\boldsymbol{k}} \\ 1 & 1 & 1 \\ -2 & -1 & 1 \end{vmatrix} = [2, -3, 1] = \mathbf{p}$$

The length of the mutual perpendicular is

$$(\mathbf{a} - \mathbf{c}) \cdot \frac{[2, -3, 1]}{\sqrt{14}} = [2, -1, 0] \cdot \frac{[2, -3, 1]}{\sqrt{14}} = 1.87 \ .$$

But the sum of the radii of the two pipes is $0.4 + 0.5 = 0.9$. Hence the pipes do not intersect.

2.2.4 The equation of a plane

There are a number of ways of specifying the equation of a plane.

1. If $\mathbf{b}$ and $\mathbf{c}$ are two non-parallel vectors (ie $\mathbf{b} \times \mathbf{c} \neq \mathbf{0}$), then the equation of the plane passing through the point $\mathbf{a}$ and parallel to the vectors $\mathbf{b}$ and $\mathbf{c}$ may be written in the form

   $$\mathbf{r} = \mathbf{a} + \lambda\mathbf{b} + \mu\mathbf{c}$$

   where λ, μ are scalar parameters. Note that $\mathbf{b}$ and $\mathbf{c}$ are free vectors, so don't have to lie in the plane (Figure 2.6(a).)

2. Figure 2.6(b) shows the plane defined by three non-collinear points $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ in the plane (note that the vectors $\mathbf{b}$ and $\mathbf{c}$ are position vectors, not free vectors as in the previous case). The equation might be written as

   $$\mathbf{r} = \mathbf{a} + \lambda(\mathbf{b} - \mathbf{a}) + \mu(\mathbf{c} - \mathbf{a})$$

3. Figure 2.6(c) illustrates another description is in terms of the unit normal to the plane $\hat{\mathbf{n}}$ and a point $\mathbf{a}$ in the plane

   $$\mathbf{r} \cdot \hat{\mathbf{n}} = \mathbf{a} \cdot \hat{\mathbf{n}} \ .$$

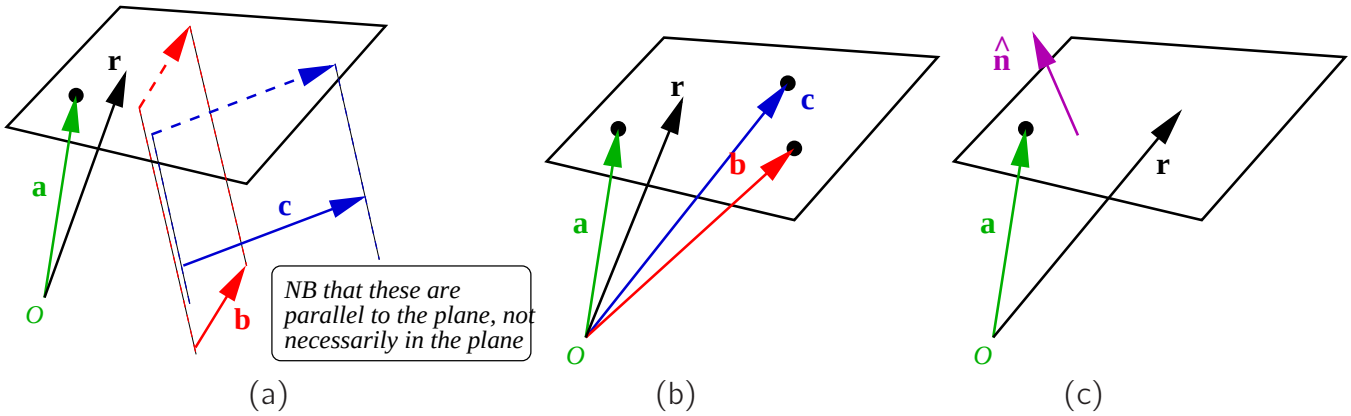


Figure 2.6: (a) Plane defined using point and two lines. (b) Plane defined using three points. (c) Plane defined using point and normal. Vector **r** is the position vector of a general point in the plane.

2.2.5 The shortest distance from a point to a plane

The shortest distance from a point **d** to the plane is along the perpendicular. Depending on how the plane is defined, this can be written as

$$D = |(\mathbf{d} - \mathbf{a}) \cdot \hat{\mathbf{n}}| \qquad \text{or} \qquad D = \frac{|(\mathbf{d} - \mathbf{a}) \cdot (\mathbf{b} \times \mathbf{c})|}{|\mathbf{b} \times \mathbf{c}|} \quad .$$

2.3 Solution of vector equations

It is sometimes required to obtain the most general vector which satisfies a given vector relationship. This is very much like obtaining the locus of a point. The best method of proceeding in such a case is as follows:

(i) Decide upon a system of three co-ordinate vectors using two non-parallel vectors appearing in the vector relationship. These might be **a**, **b** and their vector product $(\mathbf{a} \times \mathbf{b})$.

(ii) Express the unknown vector **x** as a linear combination of these vectors

$$\mathbf{x} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{a} \times \mathbf{b}$$

where λ, μ, ν are scalars to be found.

(iii) Substitute the above expression for **x** into the vector relationship to determine the constraints on λ, μ and ν for the relationship to be satisfied.

♣ Example

Q Solve the vector equation $\mathbf{x} = \mathbf{x} \times \mathbf{a} + \mathbf{b}$.

A Step (i): Set up basis vectors **a**, **b** and their vector product $\mathbf{a} \times \mathbf{b}$.
Step (ii): $\mathbf{x} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{a} \times \mathbf{b}$.

Step (iii): Bung this expression for $\mathbf{x}$ into the equation!

$$\begin{aligned}\lambda\mathbf{a}+\mu\mathbf{b}+\nu\mathbf{a}\times\mathbf{b} &= (\lambda\mathbf{a}+\mu\mathbf{b}+\nu\mathbf{a}\times\mathbf{b})\times\mathbf{a}+\mathbf{b}\\ &= \mathbf{0}+\mu(\mathbf{b}\times\mathbf{a})+\nu(\mathbf{a}\times\mathbf{b})\times\mathbf{a}+\mathbf{b}\\ &= -\nu(\mathbf{a}\cdot\mathbf{b})\mathbf{a}+(\nu a^2+1)\mathbf{b}-\mu(\mathbf{a}\times\mathbf{b})\end{aligned}$$

We have learned that any vector has a unique expression in terms of a basis set, so that the coefficients of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{a}\times\mathbf{b}$ on either side of the equation much be equal.

$$\begin{aligned}\Rightarrow\lambda &= -\nu(\mathbf{a}\cdot\mathbf{b})\\ \mu &= \nu a^2+1\\ \nu &= -\mu\end{aligned}$$

so that

$$\mu=\frac{1}{1+a^2}\qquad \nu=-\frac{1}{1+a^2}\qquad \lambda=\frac{\mathbf{a}\cdot\mathbf{b}}{1+a^2}\ .$$

So finally the solution is the single point:

$$\mathbf{x}=\frac{1}{1+a^2}\left((\mathbf{a}\cdot\mathbf{b})\mathbf{a}+\mathbf{b}-(\mathbf{a}\times\mathbf{b})\right)$$

2.4 Rotation, angular velocity/acceleration and moments

A rotation can be represented by a vector whose direction is along the axis of rotation in the sense of a r-h screw, and whose magnitude is proportional to the size of the rotation (Fig. 2.7). The same idea can be extended to the derivatives, that is, angular velocity $\boldsymbol{\omega}$ and angular acceleration $\dot{\boldsymbol{\omega}}$.

Angular accelerations arise because of a moment (or torque) on a body. In mechanics, the moment of a force $\mathbf{F}$ about a point Q is defined to have magnitude $M=Fd$, where d is the perpendicular distance between Q and the line of action L of $\mathbf{F}$.

The vector equation for moment is

$$\mathbf{M}=\mathbf{r}\times\mathbf{F}$$

where $\mathbf{r}$ is the vector from Q to *any* point on the line of action L of force $\mathbf{F}$. The resulting angular acceleration vector is in the same direction as the moment vector.

The instantaneous velocity of any point P on a rigid body undergoing pure rotation can be defined by a vector product as follows. The angular velocity vector $\boldsymbol{\omega}$ has

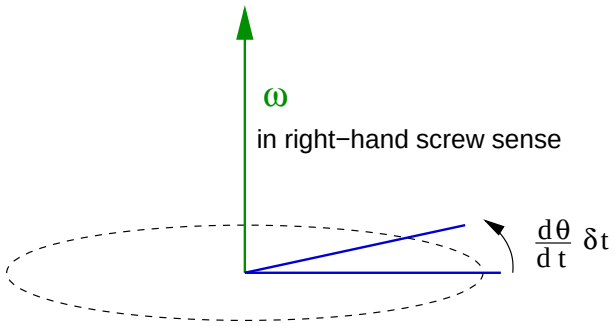


Figure 2.7: The angular velocity vector $\boldsymbol{\omega}$ is along the axis of rotation and has magnitude equal to the rate of rotation.

magnitude equal to the angular speed of rotation of the body and with direction the same as that of the r-h screw. If $\mathbf{r}$ is the vector OP, where the origin O can be taken to be any point on the axis of rotation, then the velocity $\mathbf{v}$ of P due to the rotation is given, in both magnitude and direction, by the vector product

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}.$$

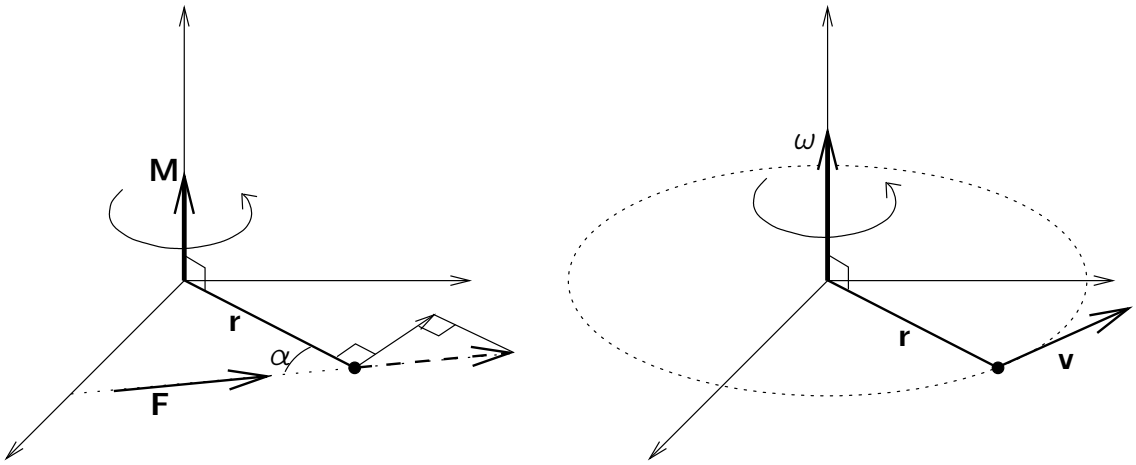


Figure 2.8:

Lecture 3

Differentiating Vector Functions of a Single Variable

Your experience of differentiation and integration has extended as far as *scalar* functions of single and multiple variables — functions like $f(x)$ and $f(x, y, t)$.

It should be no great surprise that we often wish to differentiate vector functions. For example, suppose you were driving along a wiggly road with position $\mathbf{r}(t)$ at time t. Differentiating $\mathbf{r}(t)$ wrt time should yield your velocity $\mathbf{v}(t)$, and differentiating $\mathbf{v}(t)$ should yield your acceleration. Let's see how to do this.

3.1 Differentiation of a vector

The derivative of a vector function $\mathbf{a}(p)$ of a single parameter p is

$$\mathbf{a}'(p) = \lim_{\delta p \to 0} \frac{\mathbf{a}(p + \delta p) - \mathbf{a}(p)}{\delta p} \ .$$

If we write $\mathbf{a}$ in terms of components relative to a FIXED coordinate system ($\hat{\boldsymbol{i}}, \hat{\boldsymbol{j}}, \hat{\boldsymbol{k}}$ constant)

$$\mathbf{a}(p) = a_1(p)\hat{\boldsymbol{i}} + a_2(p)\hat{\boldsymbol{j}} + a_3(p)\hat{\boldsymbol{k}}$$

then

$$\mathbf{a}'(p) = \frac{da_1}{dp}\hat{\boldsymbol{i}} + \frac{da_2}{dp}\hat{\boldsymbol{j}} + \frac{da_3}{dp}\hat{\boldsymbol{k}} \ .$$

That is, in order to differentiate a vector function, one simply differentiates each component separately. This means that all the familiar rules of differentiation apply, and they don't get altered by vector operations like scalar product and vector products.

Thus, for example:

$$\frac{d}{dp}(\mathbf{a} \times \mathbf{b}) = \frac{d\mathbf{a}}{dp} \times \mathbf{b} + \mathbf{a} \times \frac{d\mathbf{b}}{dp} \qquad \frac{d}{dp}(\mathbf{a} \cdot \mathbf{b}) = \frac{d\mathbf{a}}{dp} \cdot \mathbf{b} + \mathbf{a} \cdot \frac{d\mathbf{b}}{dp} \ .$$

Note that $d\mathbf{a}/dp$ has a different direction and a different magnitude from $\mathbf{a}$.

Likewise, as you might expect, the chain rule still applies. If $\mathbf{a} = \mathbf{a}(u)$ and $u = u(t)$, say:

$$\frac{d}{dt}\mathbf{a} = \frac{d\mathbf{a}}{du}\frac{du}{dt}$$

♣ Examples

Q A 3D vector $\mathbf{a}$ of constant magnitude is varying over time. What can you say about the direction of $\dot{\mathbf{a}}$?

A Using intuition: if only the direction is changing, then the vector must be tracing out points on the surface of a sphere. We would guess that the derivative $\dot{\mathbf{a}}$ is orthogonal to $\mathbf{a}$.

To prove this write

$$\frac{d}{dt}(\mathbf{a}\cdot\mathbf{a}) = \mathbf{a}\cdot\frac{d\mathbf{a}}{dt} + \frac{d\mathbf{a}}{dt}\cdot\mathbf{a} = 2\mathbf{a}\cdot\frac{d\mathbf{a}}{dt} \quad .$$

But $(\mathbf{a}\cdot\mathbf{a}) = a^2$ which we are told is constant. So

$$\frac{d}{dt}(\mathbf{a}\cdot\mathbf{a}) = 0 \qquad \Rightarrow 2\mathbf{a}\cdot\frac{d\mathbf{a}}{dt} = 0$$

and hence $\mathbf{a}$ and $d\mathbf{a}/dt$ must be perpendicular.

Q The position of a vehicle is $\mathbf{r}(u)$ where u is the amount of fuel consumed by some time t. Write down an expression for the acceleration.

A The velocity is

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}}{du}\frac{du}{dt}$$

$$\mathbf{a} = \frac{d}{dt}\frac{d\mathbf{r}}{dt} = \frac{d^2\mathbf{r}}{du^2}\left(\frac{du}{dt}\right)^2 + \frac{d\mathbf{r}}{du}\frac{d^2u}{dt^2}$$

3.1.1 Geometrical interpretation of vector derivatives

Let $\mathbf{r}(p)$ be a position vector tracing a space curve as some parameter p varies. The vector $\delta\mathbf{r}$ is a secant to the curve, and $\delta\mathbf{r}/\delta p$ lies in the same direction. (See Fig. 3.1.) In the limit as δp tends to zero $\delta\mathbf{r}/\delta p = d\mathbf{r}/dp$ becomes a tangent to the space curve. If the magnitude of this vector is 1 (i.e. a unit tangent), then

$|d\mathbf{r}| = dp$ so the parameter p is arc-length (metric distance). More generally, however, p will not be arc-length and we will have:

$$\frac{d\mathbf{r}}{dp} = \frac{d\mathbf{r}}{ds}\frac{ds}{dp}$$

So, the direction of the derivative is that of a **tangent to the curve**, and its magnitude is $|ds/dp|$, **the rate of change of arc length w.r.t the parameter**. Of course if that parameter p is time, the magnitude $|d\mathbf{r}/dt|$ is the speed.

♣ Example

Q Draw the curve

$$\mathbf{r} = a\cos(\frac{s}{\sqrt{a^2+h^2}})\hat{\boldsymbol{\imath}} \quad + \quad a\sin(\frac{s}{\sqrt{a^2+h^2}})\hat{\boldsymbol{\jmath}} \quad + \quad \frac{hs}{\sqrt{a^2+h^2}}\hat{\boldsymbol{k}}$$

where s is arc length and h, a are constants. Show that the tangent $d\mathbf{r}/ds$ to the curve has a constant elevation angle w.r.t the xy-plane, and determine its magnitude.

A

$$\frac{d\mathbf{r}}{ds} = -\frac{a}{\sqrt{a^2+h^2}}\sin()\,\hat{\boldsymbol{\imath}} + \frac{a}{\sqrt{a^2+h^2}}\cos()\,\hat{\boldsymbol{\jmath}} + \frac{h}{\sqrt{a^2+h^2}}\hat{\boldsymbol{k}}$$

The projection on the xy plane has magnitude $a/\sqrt{a^2+h^2}$ and in the z direction $h/\sqrt{a^2+h^2}$, so the elevation angle is a constant, $\tan^{-1}(h/a)$.

We are expecting $d\mathbf{r}/ds = 1$, and indeed

$$\sqrt{a^2\sin^2() + a^2\cos^2() + h^2}/\sqrt{a^2+h^2} = 1.$$

3.1.2 Arc length is a special parameter!

It might seem that we can be completely relaxed about saying that any old parameter p is arc length, but this is not the case. Why not? The reason is that arc length is special is that, whatever the parameter p,

$$s = \int_{p_0}^{p} \left|\frac{d\mathbf{r}}{dp}\right| dp \quad .$$

Perhaps another way to grasp the significance of this is using Pythagoras' theorem on a short piece of curve: in the limit as dx etc tend to zero,

$$ds^2 = dx^2 + dy^2 + dz^2 \ .$$

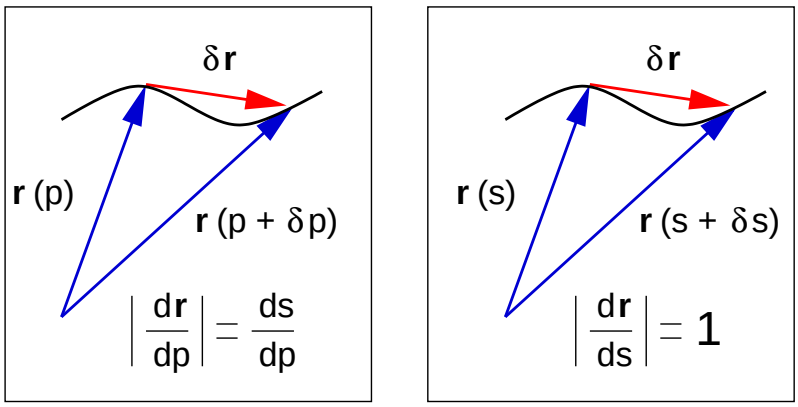


Figure 3.1: Left: $\delta\mathbf{r}$ is a secant to the curve but, in the limit as $\delta p \to 0$, becomes a tangent. Right: if the parameter is arc length s, then $|d\mathbf{r}| = ds$.

So if a curve is parameterized in terms of p

$$\frac{ds}{dp} = \sqrt{\frac{dx}{dp}^2 + \frac{dy}{dp}^2 + \frac{dz}{dp}^2} \ .$$

As an example, suppose in our earlier example we had parameterized our helix as

$$\mathbf{r} = a\cos p\hat{\boldsymbol{i}} \ + \ a\sin p\hat{\boldsymbol{j}} \ + \ hp\hat{\boldsymbol{k}}$$

It would be easy just to *say* that p was arclength, but it would not be correct because

$$\begin{aligned}\frac{ds}{dp} &= \sqrt{\frac{dx}{dp}^2 + \frac{dy}{dp}^2 + \frac{dz}{dp}^2} \\ &= \sqrt{a^2\sin^2 p + a^2\cos^2 p + h^2} = \sqrt{a^2 + h^2}\end{aligned}$$

If p really was arclength, $ds/dp = 1$. So $p/\sqrt{a^2+h^2}$ **is** arclength, not p.

3.2 Integration of a vector function

The integration of a vector function of a single scalar variable can be regarded simply as the reverse of differentiation. In other words

$$\int_{p_1}^{p_2} \frac{d\mathbf{a}(p)}{dp} dp$$

For example the integral of the acceleration vector of a point over an interval of time is equal to the change in the velocity vector during the same time interval. However, many other, more interesting and useful, types of integral are possible, especially when the vector is a function of more than one variable. This requires the introduction of the concepts of scalar and vector fields. See later!

3.3 Curves in 3 dimensions

In the examples above, parameter p has been either arc length s or time t. It doesn't have to be, but these are the main two of interest. Later we shall look at some important results when differentiating w.r.t. time, but now let use look more closely at 3D curves defined in terms of arc length, s.

Take a piece of wire, and bend it into some arbitrary non-planar curve. This is a *space curve.* We can specify a point on the wire by specifying $\mathbf{r}(s)$ as a function of distance or arc length s along the wire.

3.3.1 The Frénet-Serret relationships

We are now going to introduce a local orthogonal coordinate frame for each point s along the curve, ie one with its origin at $\mathbf{r}(s)$. To specify a coordinate frame we need three mutually perpendicular directions, and these should be *intrinsic* to the curve, not fixed in an external reference frame. The ideas were first suggested by two French mathematicians, F-J. Frénet and J. A. Serret.

1. **Tangent $\hat{\mathbf{t}}$**

   There is an obvious choice for the first direction at the point $\mathbf{r}(s)$, namely the **unit tangent $\hat{\mathbf{t}}$**. We already know that

   $$\hat{\mathbf{t}} = \frac{d\mathbf{r}(s)}{ds}$$

2. **Principal Normal $\hat{\mathbf{n}}$**

   Recall that earlier we proved that if $\mathbf{a}$ was a vector of constant magnitude that varies in direction over time then $d\mathbf{a}/dt$ was perpendicular to it. Because $\hat{\mathbf{t}}$ has constant magnitude but varies over s, $d\hat{\mathbf{t}}/ds$ must be perpendicular to $\hat{\mathbf{t}}$.

   Hence the principal normal $\hat{\mathbf{n}}$ is

   $$\frac{d\hat{\mathbf{t}}}{ds} = \kappa\hat{\mathbf{n}} : \text{ where } \kappa \geq 0 \ .$$

   κ is the **curvature**, and $\kappa = 0$ for a straight line. The plane containing $\hat{\mathbf{t}}$ and $\hat{\mathbf{n}}$ is called the **osculating plane**.

3. **The Binormal $\hat{\mathbf{b}}$**

   The local coordinate frame is completed by defining the binormal

   $$\hat{\mathbf{b}}(s) = \hat{\mathbf{t}}(s) \times \hat{\mathbf{n}}(s) \ .$$

Since $\hat{\mathbf{b}} \cdot \hat{\mathbf{t}} = 0$,

$$\frac{d\hat{\mathbf{b}}}{ds} \cdot \hat{\mathbf{t}} + \hat{\mathbf{b}} \cdot \frac{d\hat{\mathbf{t}}}{ds} = \frac{d\hat{\mathbf{b}}}{ds} \cdot \hat{\mathbf{t}} + \hat{\mathbf{b}} \cdot \kappa\hat{\mathbf{n}} = 0$$

from which

$$\frac{d\hat{\mathbf{b}}}{ds} \cdot \hat{\mathbf{t}} = 0.$$

But this means that $d\hat{\mathbf{b}}/ds$ is along the direction of $\hat{\mathbf{n}}$, or

$$\frac{d\hat{\mathbf{b}}}{ds} = -\tau(s)\hat{\mathbf{n}}(s)$$

where τ is the **torsion**, and the negative sign is a matter of convention.
Differentiating $\hat{\mathbf{n}} \cdot \hat{\mathbf{t}} = 0$ and $\hat{\mathbf{n}} \cdot \hat{\mathbf{b}} = 0$, we find

$$\frac{d\hat{\mathbf{n}}}{ds} = -\kappa(s)\hat{\mathbf{t}}(s) + \tau(s)\hat{\mathbf{b}}(s).$$

The Frénet-Serret relationships:

$$\begin{aligned} d\hat{\mathbf{t}}/ds &= \kappa\hat{\mathbf{n}} \\ d\hat{\mathbf{n}}/ds &= -\kappa(s)\hat{\mathbf{t}}(s) + \tau(s)\hat{\mathbf{b}}(s) \\ d\hat{\mathbf{b}}/ds &= -\tau(s)\hat{\mathbf{n}}(s) \end{aligned}$$

♣ Example

Q Derive $\kappa(s)$ and $\tau(s)$ for the helix

$$\mathbf{r}(s) = a\cos\left(\frac{s}{\beta}\right)\hat{\boldsymbol{\imath}} + a\sin\left(\frac{s}{\beta}\right)\hat{\boldsymbol{\jmath}} + h\left(\frac{s}{\beta}\right)\hat{\boldsymbol{k}}; \qquad \beta = \sqrt{a^2 + h^2}$$

and comment on their values.

A We found the unit tangent earlier as

$$\hat{\mathbf{t}} = \frac{d\mathbf{r}}{ds} = \left[-\frac{a}{\beta}\sin\left(\frac{s}{\beta}\right), \quad \frac{a}{\beta}\cos\left(\frac{s}{\beta}\right), \quad \frac{h}{\beta}\right] .$$

Differentiation gives

$$\kappa\hat{\mathbf{n}} = \frac{d\hat{\mathbf{t}}}{ds} = \left[-\frac{a}{\beta^2}\cos\left(\frac{s}{\beta}\right), \quad -\frac{a}{\beta^2}\sin\left(\frac{s}{\beta}\right), \quad 0\right]$$

Curvature is always positive, so

$$\kappa = \frac{a}{\beta^2} \qquad \hat{\mathbf{n}} = \left[-\cos\left(\frac{s}{\beta}\right), \quad -\sin\left(\frac{s}{\beta}\right), \quad 0 \right] .$$

So the curvature is constant, and the normal is parallel to the xy-plane. Now use

$$\hat{\mathbf{b}} = \hat{\mathbf{t}} \times \hat{\mathbf{n}} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ (-a/\beta)S & (a/\beta)C & (h/\beta) \\ -C & -S & 0 \end{vmatrix} = \left[\frac{h}{\beta}\sin\left(\frac{s}{\beta}\right), \quad -\frac{h}{\beta}\cos\left(\frac{s}{\beta}\right), \quad \frac{a}{\beta} \right]$$

and differentiate $\hat{\mathbf{b}}$ to find an expression for the torsion

$$\frac{d\hat{\mathbf{b}}}{ds} = \left[\frac{h}{\beta^2}\cos\left(\frac{s}{\beta}\right), \quad \frac{h}{\beta^2}\sin\left(\frac{s}{\beta}\right), \quad 0 \right] = \frac{-h}{\beta^2}\hat{\mathbf{n}}$$

so the torsion is

$$\tau = \frac{h}{\beta^2}$$

again a constant.

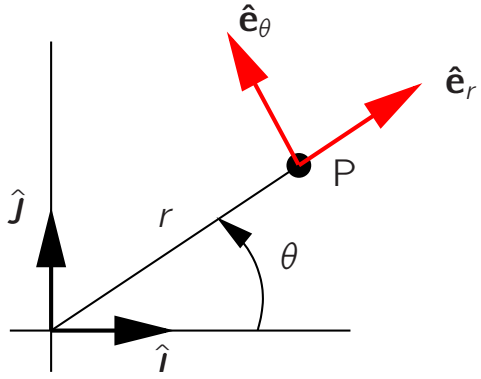
3.4 Radial and tangential components in plane polars

In plane polar coordinates, the radius vector of any point P is given by

$$\begin{aligned} \mathbf{r} &= r\cos\theta\hat{\imath} + r\sin\theta\hat{\jmath} \\ &= r\hat{\mathbf{e}}_r \end{aligned}$$

where we have introduced the unit radial vector

$$\hat{\mathbf{e}}_r = \cos\theta\hat{\imath} + \sin\theta\hat{\jmath} \quad .$$



The other "natural" (we'll see why in a later lecture) unit vector in plane polars is orthogonal to $\hat{\mathbf{e}}_r$ and is

$$\hat{\mathbf{e}}_\theta = -\sin\theta\hat{\imath} + \cos\theta\hat{\jmath}$$

so that $\hat{\mathbf{e}}_r \cdot \hat{\mathbf{e}}_r = \hat{\mathbf{e}}_\theta \cdot \hat{\mathbf{e}}_\theta = 1$ and $\hat{\mathbf{e}}_r \cdot \hat{\mathbf{e}}_\theta = 0$.

Now suppose P is moving so that $\mathbf{r}$ is a function of time t. Its velocity is

$$\begin{aligned}\dot{\mathbf{r}} = \frac{d}{dt}(r\hat{\mathbf{e}}_r) &= \frac{dr}{dt}\hat{\mathbf{e}}_r + r\frac{d\hat{\mathbf{e}}_r}{dt} \\ &= \frac{dr}{dt}\hat{\mathbf{e}}_r + r\frac{d\theta}{dt}(-\sin\theta\hat{\boldsymbol{i}} + \cos\theta\hat{\boldsymbol{j}}) \\ &= \frac{dr}{dt}\hat{\mathbf{e}}_r + r\frac{d\theta}{dt}\hat{\mathbf{e}}_\theta \\ &= \text{radial} + \text{tangential}\end{aligned}$$

The radial and tangential components of velocity of P are therefore dr/dt and $rd\theta/dt$, respectively.

Differentiating a second time gives the acceleration of P

$$\begin{aligned}\ddot{\mathbf{r}} &= \frac{d^2r}{dt^2}\hat{\mathbf{e}}_r + \frac{dr}{dt}\frac{d\theta}{dt}\hat{\mathbf{e}}_\theta + \frac{dr}{dt}\frac{d\theta}{dt}\hat{\mathbf{e}}_\theta + r\frac{d^2\theta}{dt^2}\hat{\mathbf{e}}_\theta - r\frac{d\theta}{dt}\frac{d\theta}{dt}\hat{\mathbf{e}}_r \\ &= \left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right]\hat{\mathbf{e}}_r + \left[2\frac{dr}{dt}\frac{d\theta}{dt} + r\frac{d^2\theta}{dt^2}\right]\hat{\mathbf{e}}_\theta\end{aligned}$$

3.5 Rotating systems

Consider a body which is rotating with constant angular velocity $\boldsymbol{\omega}$ about some axis passing through the origin. Assume the origin is fixed, and that we are sitting in a fixed coordinate system $Oxyz$.

If $\boldsymbol{\rho}$ is a vector of constant magnitude and constant direction in the rotating system, then its representation $\mathbf{r}$ in the fixed system must be a function of t.

$$\mathbf{r}(t) = \mathrm{R}(t)\boldsymbol{\rho}$$

At any instant as observed in the fixed system

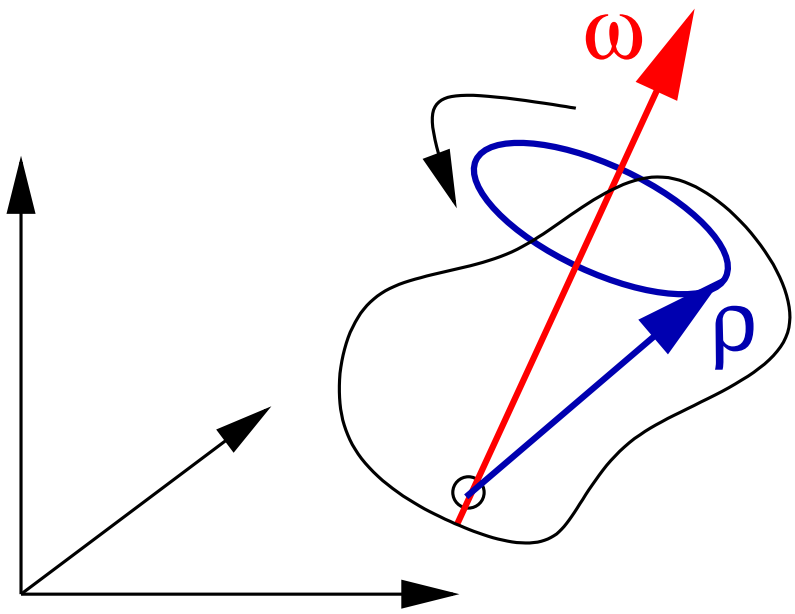
$$\frac{d\mathbf{r}}{dt} = \dot{\mathrm{R}}\boldsymbol{\rho} + \mathrm{R}\dot{\boldsymbol{\rho}}$$

but the second term is zero since we assumed $\boldsymbol{\rho}$ to be constant so we have

$$\frac{d\mathbf{r}}{dt} = \dot{\mathrm{R}}\mathrm{R}^{\top}\mathbf{r}$$

Note that:

- $d\mathbf{r}/dt$ will have fixed magnitude;
- $d\mathbf{r}/dt$ will always be perpendicular to the axis of rotation;
- $d\mathbf{r}/dt$ will vary in direction within those constraints;
- $\mathbf{r}(t)$ will move in a plane in the fixed system.



Now let's consider the term $\dot{\mathsf{R}}\mathsf{R}^\top$. First, note that $\mathsf{R}\mathsf{R}^\top = \mathsf{I}$ (the identity), so differentiating both sides yields

$$\begin{aligned} \dot{\mathsf{R}}\mathsf{R}^\top + \mathsf{R}\dot{\mathsf{R}}^\top &= 0 \\ \dot{\mathsf{R}}\mathsf{R}^\top &= -\mathsf{R}\dot{\mathsf{R}}^\top \end{aligned}$$

Thus $\dot{\mathsf{R}}\mathsf{R}^\top$ is anti-symmetric:

$$\dot{\mathsf{R}}\mathsf{R}^\top = \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}$$

Now you can verify for yourself that application of a matrix of this form to an arbitrary vector has precisely the same effect as the cross product operator, $\boldsymbol{\omega}\times$, where $\boldsymbol{\omega} = [xyz]^\top$. Loh-and-behold, we then we have

$$\dot{\mathbf{r}} = \boldsymbol{\omega} \times \mathbf{r}$$

matching the equation at the end of lecture 2, $\mathbf{v} = \boldsymbol{\omega}\times\mathbf{r}$, as we would hope/expect.

3.5.1 Rotation: Part 2

Now suppose $\boldsymbol{\rho}$ is the position vector of a point P which **moves** in the rotating frame. There will be two contributions to motion with respect to the fixed frame, one due to its motion within the rotating frame, and one due to the rotation itself. So, returning to the equations we derived earlier:

$$\mathbf{r}(t) = \mathsf{R}(t)\boldsymbol{\rho}(t)$$

and the instantaenous differential with respect to time:

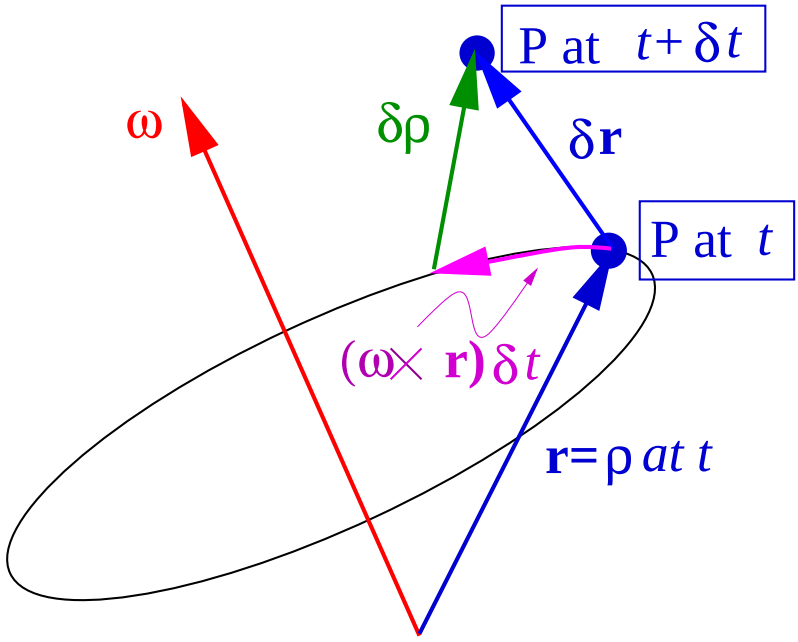
$$\frac{d\mathbf{r}}{dt} = \dot{\mathsf{R}}\boldsymbol{\rho} + \mathsf{R}\dot{\boldsymbol{\rho}} = \dot{\mathsf{R}}\mathsf{R}^\top\mathbf{r} + \mathsf{R}\dot{\boldsymbol{\rho}}$$

Now $\boldsymbol{\rho}$ is not constant, so its differential is not zero; hence rewriting this last equations we have that

The **instantaneous velocity** of P in the fixed frame is

$$\frac{d\mathbf{r}}{dt} = \mathsf{R}\dot{\boldsymbol{\rho}} + \boldsymbol{\omega} \times \mathbf{r}$$

The second term of course, is the contribution from the rotating frame which we saw previously. The first is the linear velocity measured in the rotating frame $\dot{\boldsymbol{\rho}}$, referred to the fixed frame (via the rotation matrix R which aligns the two frames)



3.5.2 Rotation 3: Instantaneous acceleration

Our previous result is a general one relating the time derivatives of any vector in rotating and non-rotating frames. Let us now consider the second differential:

$$\ddot{\mathbf{r}} = \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times \dot{\mathbf{r}} + \dot{\mathrm{R}}\dot{\boldsymbol{\rho}} + \mathrm{R}\ddot{\boldsymbol{\rho}}$$

We shall assume that the angular acceleration is zero, which kills off the first term, and so now, substituting for $\dot{\mathbf{r}}$ we have

$$\begin{aligned}
\ddot{\mathbf{r}} &= \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r} + \mathrm{R}\dot{\boldsymbol{\rho}}) + \dot{\mathrm{R}}\dot{\boldsymbol{\rho}} + \mathrm{R}\ddot{\boldsymbol{\rho}} \\
&= \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + \boldsymbol{\omega} \times \mathrm{R}\dot{\boldsymbol{\rho}} + \dot{\mathrm{R}}\dot{\boldsymbol{\rho}} + \mathrm{R}\ddot{\boldsymbol{\rho}} \\
&= \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + \boldsymbol{\omega} \times \mathrm{R}\dot{\boldsymbol{\rho}} + \dot{\mathrm{R}}(\mathrm{R}^{\top}\mathrm{R})\dot{\boldsymbol{\rho}} + \mathrm{R}\ddot{\boldsymbol{\rho}} \\
&= \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + 2\boldsymbol{\omega} \times (\mathrm{R}\dot{\boldsymbol{\rho}}) + \mathrm{R}\ddot{\boldsymbol{\rho}}
\end{aligned}$$

The **instantaneous acceleration** is therefore

$$\ddot{\mathbf{r}} = \mathrm{R}\ddot{\boldsymbol{\rho}} + 2\boldsymbol{\omega} \times (\mathrm{R}\dot{\boldsymbol{\rho}}) + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})$$

- The first term is the acceleration of the point P in the rotating frame measured in the rotating frame, but referred to the fixed frame by the rotation R
- The last term is the centripetal acceleration to due to the rotation. (Yes! Its magnitude is $\omega^2 r$ and its direction is that of $-\mathbf{r}$. Check it out.)

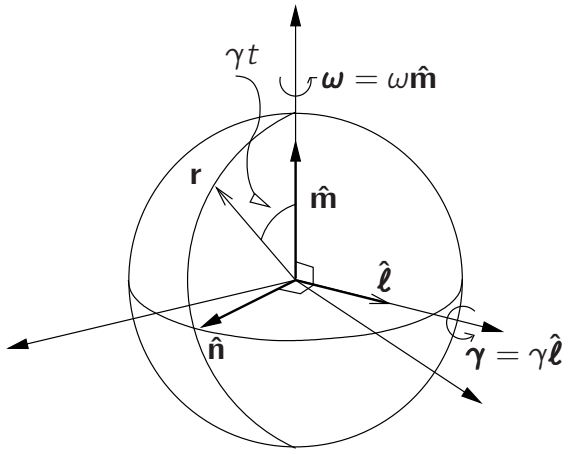


Figure 3.2: Coriolis example.

- The middle term is an extra term which arises because of the velocity of P in the rotating frame. It is known as the **Coriolis acceleration**, named after the French engineer who first identified it.

Because of the rotation of the earth, the Coriolis acceleration is of great importance in meteorology and accounts for the occurrence of high pressure anticyclones and low pressure cyclones in the northern hemisphere, in which the Coriolis acceleration is produced by a pressure gradient. It is also a very important component of the acceleration (hence the force exerted) by a rapidly moving robot arm, whose links whirl rapidly about rotary joints.

♣ Example

Q Find the instantaneous acceleration of a projectile fired along a line of longitude (with angular velocity of $\boldsymbol{\gamma}$ constant relative to the sphere) if the sphere is rotating with angular velocity $\boldsymbol{\omega}$.

A Consider a coordinate frame defined by mutually orthogonal unit vectors, $\hat{\boldsymbol{\ell}}, \hat{\mathbf{m}}$ and $\hat{\mathbf{n}}$, as shown in Fig. 3.2. We shall assume, without loss of generality, that the fixed and rotating frames are instantaneously aligned at the moment shown in the diagram, so that $\mathsf{R} = \mathsf{I}$, the identity, and hence $\mathbf{r} = \boldsymbol{\rho}$.

In the rotating frame

$$\dot{\boldsymbol{\rho}} = \boldsymbol{\gamma} \times \boldsymbol{\rho} \qquad \text{and} \qquad \ddot{\boldsymbol{\rho}} = \boldsymbol{\gamma} \times \dot{\boldsymbol{\rho}} = \boldsymbol{\gamma} \times (\boldsymbol{\gamma} \times \boldsymbol{\rho})$$

So the in the fixed reference frame, because these two frames are instantaneously aligned

$$\ddot{\mathbf{r}} \;=\; \boldsymbol{\gamma} \times (\boldsymbol{\gamma} \times \boldsymbol{\rho}) \;\;+\;\; 2\boldsymbol{\omega} \times (\boldsymbol{\gamma} \times \boldsymbol{\rho}) \;\;+\;\; \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \;.$$

The first term is the centripetal acceleration due to the projectile moving around the sphere — which it does because of the gravitational force. The

last term is the centripetal acceleration resulting from the rotation of the sphere. The middle term is the Coriolis acceleration.

Using Fig. 3.2, at some instant t

$$\mathbf{r}(t) = \boldsymbol{\rho}(t) = r\cos(\gamma t)\hat{\mathbf{m}} \quad + \quad r\sin(\gamma t)\hat{\mathbf{n}}$$

and

$$\boldsymbol{\gamma} = \gamma\hat{\boldsymbol{\ell}}$$

Then

$$\boldsymbol{\gamma} \times (\boldsymbol{\gamma} \times \boldsymbol{\rho}) = (\boldsymbol{\gamma} \cdot \boldsymbol{\rho})\boldsymbol{\gamma} - \gamma^2\boldsymbol{\rho} = -\gamma^2\boldsymbol{\rho} = -\gamma^2\mathbf{r},$$

Check the direction — the negative sign means it points *towards* the centre of the sphere, which is as expected.

Likewise the last term can be obtained as

$$\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) = -\omega^2 r\sin(\gamma t)\hat{\mathbf{n}}$$

Note that it is perpendicular to the axis of rotation $\hat{\mathbf{m}}$, and because of the minus sign, directed towards the axis)

The Coriolis term is derived as:

$$\begin{aligned} 2\boldsymbol{\omega} \times \dot{\boldsymbol{\rho}} &= 2\boldsymbol{\omega} \times (\boldsymbol{\gamma} \times \boldsymbol{\rho}) \\ &= 2\begin{bmatrix}0\\ \omega\\ 0\end{bmatrix} \times \left(\begin{bmatrix}\gamma\\ 0\\ 0\end{bmatrix} \times \begin{bmatrix}0\\ r\cos\gamma t\\ r\sin\gamma t\end{bmatrix}\right) \\ &= 2\omega\gamma r\cos\gamma t\hat{\boldsymbol{\ell}} \end{aligned}$$

Instead of a projectile, now consider a rocket on rails which stretch north from the equator. As the rocket travels north it experiences the Coriolis force (exerted by the rails):

$$\begin{array}{cccc} 2 & \gamma & \omega & R\cos\gamma t \quad \hat{\boldsymbol{\ell}} \\ +\text{ve} & \text{-ve} & +\text{ve} & +\text{ve} \end{array}$$

Hence the coriolis force is in the direction opposed to $\hat{\boldsymbol{\ell}}$ (i.e. in the opposite direction to the earth's rotation). In the absence of the rails (or atmosphere) the rocket's tangetial speed (relative to the surface of the earth) is *greater* than the speed of the surface of the earth underneath it (since the radius of successive lines of latitude decreases) so it would (to an observer on the earth) appear to deflect to the east. The rails provide a coriolis force keeping it on the same meridian.

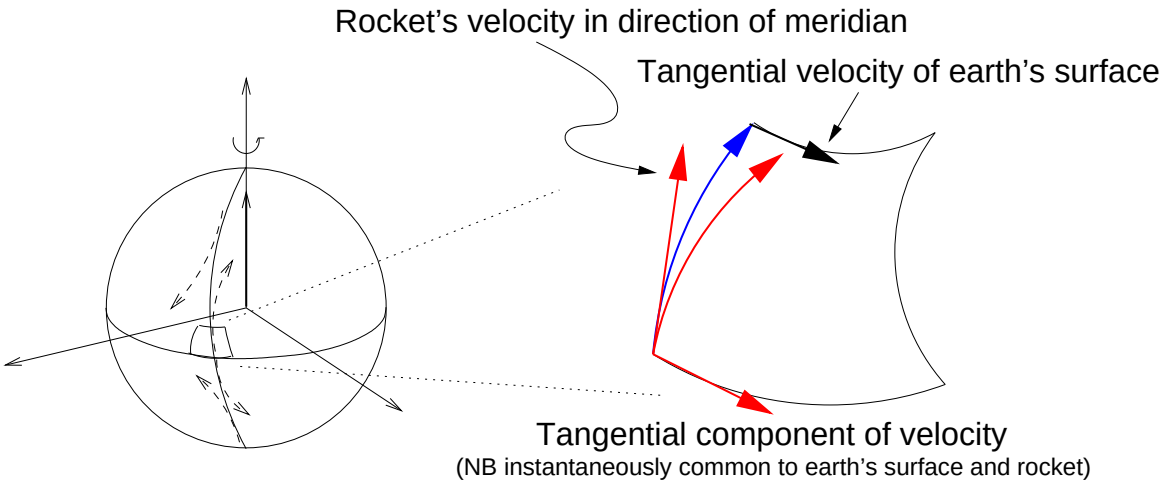

Rocket's velocity in direction of meridian
Tangential velocity of earth's surface
Tangential component of velocity
(NB instantaneously common to earth's surface and rocket)

Figure 3.3: Rocket example

Figure 3.4: Coriolis effect giving rise to weather systems

Lecture 4

Line, Surface and Volume Integrals. Curvilinear coordinates.

We started off the course being concerned with individual vectors **a**, **b**, **c**, and so on.

We went on to consider how single vectors vary over time or over some other parameter such as arc length.

In much of the rest of the course, we will be concerned with scalars and vectors which are defined over regions in space — scalar and vector *fields*

In this lecture we introduce line, surface and volume integrals, and consider how these are defined in non-Cartesian, *curvilinear coordinates*

4.1 Scalar and vector fields

When a scalar function $u(\mathbf{r})$ is determined or defined at each position $\mathbf{r}$ in some region, we say that u is a **scalar field** in that region.

Similarly, if a vector function $\mathbf{v}(\mathbf{r})$ is defined at each point, then $\mathbf{v}$ is a **vector field** in that region. As you will see, in **field theory** our aim is to derive statements about the bulk properties of scalar and vector fields, rather than to deal with individual scalars or vectors.

Familiar examples of each are shown in figure 4.1.

In Lecture 1 we worked out the force $\mathbf{F}(\mathbf{r})$ on a charge Q arising from a number of charges q_i. The electric field is $\mathbf{F}/Q$, so

$$\mathbf{E}(\mathbf{r}) = \sum_{i=1}^{N} K \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|^3}(\mathbf{r} - \mathbf{r}_i) \quad . \qquad (K = \frac{1}{4\pi\epsilon_r\epsilon_0})$$

For example; you could work out the velocity field, in plane polars, at any point on

(a)

(b)

Figure 4.1: Examples of (a) a scalar field (pressure); (b) avector field (wind velocity)

a wheel spinning about its axis

$$\mathbf{v}(\mathbf{r}) = \boldsymbol{\omega} \times \mathbf{r}$$

or the fluid flow field around a wing.

If the fields are independent of time, they are said to be *steady*. Of course, most vector fields of practical interest in engineering science are not steady, and some are unpredictable.

Let us first consider how to perform a variety of types of integration in vector and scalar fields.

4.2 Line integrals through fields

Line integrals are concerned with measuring the integrated interaction with a field as you move through it on some defined path. Eg, given a map showing the pollution density field in Oxford, you may wish to work out how much pollution you breathe in when cycling from college to the Department via different routes.

First recall the definition of an integral for a scalar function $f(x)$ of a single scalar variable x. One assumes a set of n samples $f_i = f(x_i)$ spaced by δx_i. One forms the limit of the sum of the products $f(x_i)\delta x_i$ as the number of samples tends to infinity

$$\int f(x)dx = \lim_{\substack{n \to \infty \\ \delta x_i \to 0}} \sum_{i=1}^{n} f_i \delta x_i \quad .$$

For a smooth function, it is irrelevant how the function is subdivided.

4.2.1 Vector line integrals

In a vector line integral, the path L along which the integral is to be evaluated is split into a large number of *vector* segments $\delta\mathbf{r}_i$. Each line segment is then

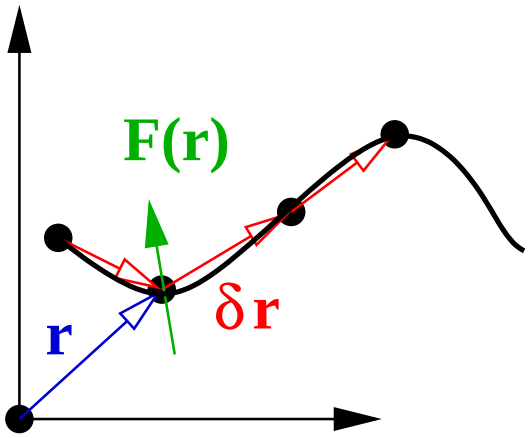


Figure 4.2: Line integral. In the diagram $\mathbf{F}(\mathbf{r})$ is a vector field, but it could be replaced with scalar field $U(\mathbf{r})$.

multiplied by the quantity associated with that point in space, the products are then summed and the limit taken as the lengths of the segments tend to zero.

There are three types of integral we have to think about, depending on the nature of the product:

1. Integrand $U(\mathbf{r})$ is a scalar field, hence the integral is a vector.

$$\mathbf{I} = \int_L U(\mathbf{r}) d\mathbf{r} \qquad \left(= \lim_{\delta \mathbf{r}_i \to 0} \sum_i U_i \boldsymbol{\delta} \mathbf{r}_i . \right)$$

2. Integrand $\mathbf{a}(\mathbf{r})$ is a vector field dotted with $d\mathbf{r}$ hence the integral is a scalar:

$$I = \int_L \mathbf{a}(\mathbf{r}) \cdot d\mathbf{r} \qquad \left(= \lim_{\delta \mathbf{r}_i \to 0} \sum_i \mathbf{a}_i \cdot \boldsymbol{\delta} \mathbf{r}_i . \right)$$

3. Integrand $\mathbf{a}(\mathbf{r})$ is a vector field crossed with $d\mathbf{r}$ hence vector result.

$$\mathbf{I} = \int_L \mathbf{a}(\mathbf{r}) \times d\mathbf{r} \qquad \left(= \lim_{\delta \mathbf{r}_i \to 0} \sum_i \mathbf{a}_i \times \boldsymbol{\delta} \mathbf{r}_i . \right)$$

Note immediately that unlike an integral in a single scalar variable, there are many paths L from start point $\mathbf{r}_A$ to end point $\mathbf{r}_B$, and the integral will in general depend on the path taken.

Physical examples of line integrals

- The total work done by a force $\mathbf{F}$ as it moves a point from A to B along a given path C is given by a line integral of type 2 above. If the force acts

at point $\mathbf{r}$ and the instantaneous displacement along curve C is $d\mathbf{r}$ then the infinitessimal work done is $dW = \mathbf{F}.d\mathbf{r}$, and so the total work done traversing the path is

$$W_C = \int_C \mathbf{F}.d\mathbf{r}$$

- Ampère's law relating magnetic field $\mathbf{B}$ to linked current can be written as

$$\oint_C \mathbf{B}.d\mathbf{r} = \mu_0 I$$

where I is the current enclosed by (closed) path C.

- The force on an element of wire carrying current I, placed in a magnetic field of strength $\mathbf{B}$, is $d\mathbf{F} = I d\mathbf{r} \times \mathbf{B}$. So if a loop this wire C is placed in the field then the total force will be and integral of type 3 above:

$$\mathbf{F} = I \oint_C d\mathbf{r} \times \mathbf{B}$$

Note that the expressions above are beautifully compact in vector notation, and are all independent of coordinate system. Of course when evaluating them we need to choose a coordinate system: often this is the standard Cartesian coordinate system (as in the worked examples below), but need not be, as we shall see in section 4.6.

♣ Examples

Q1 An example in the xy-plane. A force $\mathbf{F} = x^2y\hat{\boldsymbol{\imath}} + xy^2\hat{\boldsymbol{\jmath}}$ acts on a body as it moves between $(0,0)$ and $(1,1)$.

Determine the work done when the path is

1. along the line $y = x$.
2. along the curve $y = x^n, n > 0$.
3. along the x axis to the point $(1,0)$ and then along the line $x = 1$.

A1 This is an example of the "type 2" line integral. In planar Cartesians, $d\mathbf{r} = \hat{\boldsymbol{\imath}}dx + \hat{\boldsymbol{\jmath}}dy$. Then the work done is

$$\int_L \mathbf{F} \cdot d\mathbf{r} = \int_L (x^2y dx + xy^2 dy) \ .$$

1. For the path $y = x$ we find that $dy = dx$. So it is easiest to convert all y references to x.

$$\int_{(0,0)}^{(1,1)} (x^2y dx + xy^2 dy) = \int_{x=0}^{x=1} (x^2x dx + xx^2 dx) = \int_{x=0}^{x=1} 2x^3 dx = \left[x^4/2\right]_{x=0}^{x=1} = 1/2 \ .$$

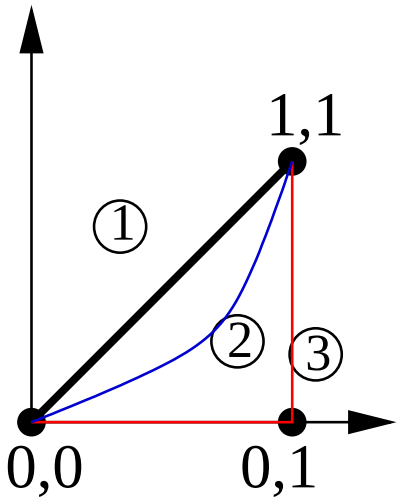


Figure 4.3: Line integral taken along three difference paths

2. For the path $y = x^n$ we find that $dy = nx^{n-1}dx$, so again it is easiest to convert all y references to x.

$$\begin{aligned}\int_{(0,0)}^{(1,1)} (x^2 y\,dx + xy^2 dy) &= \int_{x=0}^{x=1} (x^{n+2}dx + nx^{n-1}.x.x^{2n}dx) \\ &= \int_{x=0}^{x=1} (x^{n+2}dx + nx^{3n}dx) \\ &= \frac{1}{n+3} + \frac{n}{3n+1}\end{aligned}$$

3. This path is not smooth, so break it into two. Along the first section, $y = 0$ and $dy = 0$, and on the second $x = 1$ and $dx = 0$, so

$$\int_A^B (x^2 y\,dx + xy^2 dy) = \int_{x=0}^{x=1} (x^2 0 dx) + \int_{y=0}^{y=1} 1y^2 dy = 0 + \left[y^3/3\right|_{y=0}^{y=1} = 1/3\,.$$

So in general the integral depends on the path taken. Notice that answer (1) is the same as answer (2) when $n = 1$, and that answer (3) is the limiting value of answer (2) as $n \to \infty$.

Q2 Repeat part (2) using the Force $\mathbf{F} = xy^2\hat{\boldsymbol{\imath}} + x^2 y\hat{\boldsymbol{\jmath}}$.

A2 For the path $y = x^n$ we find that $dy = nx^{n-1}dx$, so

$$\begin{aligned}\int_{(0,0)}^{(1,1)}(y^2xdx + yx^2dy) &= \int_{x=0}^{x=1}(x^{2n+1}dx + nx^{n-1}.x^2.x^n dx)\\ &= \int_{x=0}^{x=1}(x^{2n+1}dx + nx^{2n+1}dx)\\ &= \frac{1}{2n+2} + \frac{n}{2n+2}\\ &= \frac{1}{2} \quad \text{independent of } n\end{aligned}$$

4.3 Line integrals in Conservative fields

In the second example, the line integral has the same value for the whole range of paths. In fact it is wholly independent of path. This is easy to see if we write $g(x, y) = x^2y^2/2$. Then using the definition of the perfect differential

$$dg = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy$$

we find that

$$\begin{aligned}\int_A^B(y^2xdx + yx^2dy) &= \int_A^B dg\\ &= g_B - g_A\end{aligned}$$

which depends solely on the value of g at the start and end points, and not at all on the path used to get from A to B. Such a vector field is called **conservative**.

One sort of line integral performs the integration around a complete loop and is denoted with a ring. If **E** is a conservative field, determine the value of

$$\oint \mathbf{E}\cdot d\mathbf{r}\ .$$

In electrostatics, if **E** is the electric field then the potential function is

$$\phi = -\int \mathbf{E}\cdot d\mathbf{r}\ .$$

Do you think **E** is conservative?

4.3.1 A note on line integrals defined in terms of arc length

Line integrals are often defined in terms of scalar arc length. They don't appear to involve vectors (but actually they are another form of type 2 defined earlier).

The integrals usually appear as follows

$$I = \int_L F(x, y, z) ds$$

and most often the path L is along a curve defined parametrically as $x = x(p)$, $y = y(p)$, $z = z(p)$ where p is some parameter. Convert the function to $F(p)$, writing

$$I = \int_{p_{\text{start}}}^{p_{\text{end}}} F(p) \frac{ds}{dp}\, dp$$

where

$$\frac{ds}{dp} = \left[\left(\frac{dx}{dp}\right)^2 + \left(\frac{dy}{dp}\right)^2 + \left(\frac{dz}{dp}\right)^2 \right]^{1/2} .$$

Note that the parameter p could be arc-length s itself, in which case $ds/dp = 1$ of course! Another possibility is that the parameter p is x — that is we are told $y = y(x)$ and $z = z(x)$. Then

$$I = \int_{x_{\text{start}}}^{x_{\text{end}}} F(x) \left[1 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2 \right]^{1/2} dx \ .$$

4.4 Surface integrals

These can be defined by analogy with line integrals.

The surface S over which the integral is to be evaluated is now divided into infinitesimal vector elements of area $d\mathbf{S}$, the direction of the vector $d\mathbf{S}$ representing the direction of the surface normal and its magnitude representing the area of the element.

Again there are three possibilities:

- $\int_S U d\mathbf{S}$ — scalar field U; vector integral.
- $\int_S \mathbf{a} \cdot d\mathbf{S}$ — vector field $\mathbf{a}$; scalar integral.
- $\int_S \mathbf{a} \times d\mathbf{S}$ — vector field $\mathbf{a}$; vector integral.

(in addition, of course, to the purely scalar form, $\int_S U dS$).

Physical example of surface integral

- Physical examples of surface integrals with vectors often involve the idea of *flux* of a vector field through a surface, $\int_S \mathbf{a}.d\mathbf{S}$ For example the mass of fluid crossing a surface S in time dt is $dM = \rho\mathbf{v}.d\mathbf{S}dt$ where $\rho(\mathbf{r})$ is the fluid density and $\mathbf{v}(\mathbf{r})$ is the fluid velocity. The total mass flux can be expressed as a surface integral:

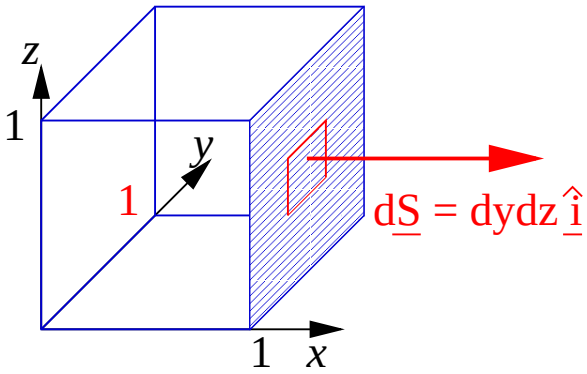
$$\Phi_M = \int_S \rho(r)\mathbf{v}(\mathbf{r}).d\mathbf{S}$$

Again, though this expression is coordinate free, we evaluate an example below using Cartesians. Note, however, that in some problems, symmetry may lead us to a different more natural coordinate system.

♣ **Example**

Evaluate $\int \mathbf{F}\cdot d\mathbf{S}$ over the $x = 1$ side of the cube shown in the figure when $\mathbf{F} = y\hat{\boldsymbol{\imath}} + z\hat{\boldsymbol{\jmath}} + x\hat{\boldsymbol{k}}$.
$d\mathbf{S}$ is perpendicular to the surface. Its $\pm$ direction actually depends on the nature of the problem. More often than not, the surface will enclose a volume, and the surface direction is taken as everywhere emanating from the interior.
Hence for the $x = 1$ face of the cube

$$d\mathbf{S} = dydz\hat{\boldsymbol{\imath}}$$



and

$$\int \mathbf{F}\cdot d\mathbf{S} = \int\int ydydz$$
$$= \frac{1}{2}y^2\Big|_0^1 z\Big|_0^1 = \frac{1}{2} \quad .$$

4.5 Volume integrals

The definition of the volume integral is again taken as the limit of a sum of products as the size of the volume element tends to zero. One obvious difference though is that the element of volume is a scalar (how could you define a direction with an infinitesimal volume element?). The possibilities are:

- $\int_V U(\mathbf{r})dV$ — scalar field; scalar integral.
- $\int_V \mathbf{a}dV$ — vector field; vector integral.

You have covered these (more or less) in your first year course, so not much more to say here. The next section considers these again in the context of a change of coordinates.

4.6 Changing variables: curvilinear coordinates

Up to now we have been concerned with Cartesian coordinates x, y, z with coordinate axes $\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}$. When performing a line integral in Cartesian coordinates, you write

$$\mathbf{r} = x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}} \qquad \text{and} \qquad d\mathbf{r} = dx\hat{\boldsymbol{\imath}} + dy\hat{\boldsymbol{\jmath}} + dz\hat{\boldsymbol{k}}$$

and can be sure that length scales are properly handled because – as we saw in Lecture 3 –

$$|d\mathbf{r}| = ds = \sqrt{dx^2 + dy^2 + dz^2}\ .$$

The reason for using the basis $\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}$ rather than any other orthonormal basis set is that $\hat{\boldsymbol{\imath}}$ represents a direction in which x is increasing while the other two coordinates remain constant (and likewise for $\hat{\boldsymbol{\jmath}}$ and $\hat{\boldsymbol{k}}$ with y and z respectively), simplifying the representation and resulting mathematics.

Often the symmetry of the problem strongly hints at using another coordinate system:

- likely to be plane, cylindrical, or spherical polars,
- but can be something more exotic

The general name for any different "u, v, w" coordinate system is a **curvilinear coordinate system**. We will see that the idea hinted at above – of defining a basis set by considering directions in which only one coordinate is (instantaneously) increasing – provides the approriate generalisation.

We begin by discussing common special cases: cylindrical polars and spherical polars, and conclude with a more general formulation.

4.6.1 Cylindrical polar coordinates

As shown in figure 4.4 a point in space P having cartesian coordinates x, y, z can be expressed in terms of cylindrical polar coordinates, r, ϕ, z as follows:

$$\begin{aligned}\mathbf{r} &= x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}} \\ &= r\cos\phi\hat{\boldsymbol{\imath}} + r\sin\phi\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}\end{aligned}$$

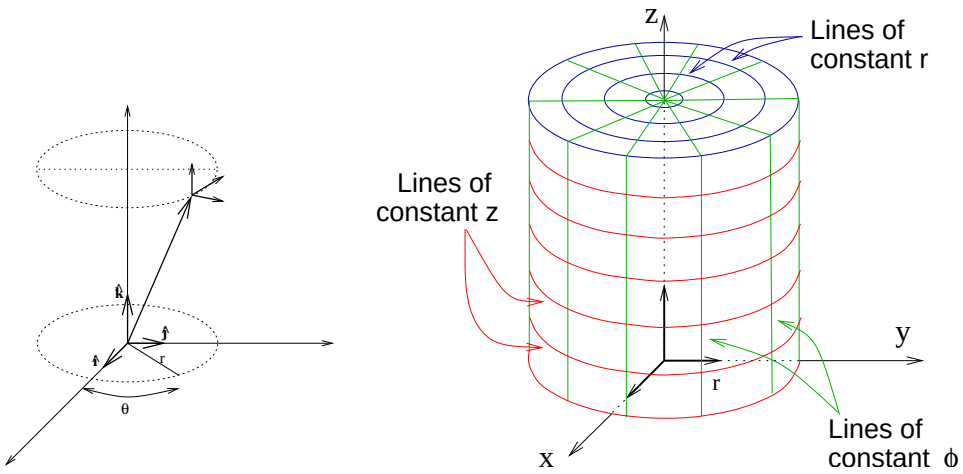


Figure 4.4: Cylindrical polars: (a) coordinate definition; (b) "iso" lines in r, ϕ and z.

Note that, by definition, $\frac{\partial \mathbf{r}}{\partial r}$ represents a direction in which (instantaneously) r is changing while the other two coordinates stay constant. That is, it is tangent to lines of constant ϕ and z. Likewise for $\frac{\partial \mathbf{r}}{\partial \phi}$ and $\frac{\partial \mathbf{r}}{\partial z}$, Thus the vectors:

$$\begin{aligned}
\mathbf{e}_r &= \frac{\partial \mathbf{r}}{\partial r} = \cos\phi\hat{\boldsymbol{\imath}} + \sin\phi\hat{\boldsymbol{\jmath}} \\
\mathbf{e}_\phi &= \frac{\partial \mathbf{r}}{\partial \phi} = -r\sin\phi\hat{\boldsymbol{\imath}} + r\cos\phi\hat{\boldsymbol{\jmath}} \\
\mathbf{e}_z &= \frac{\partial \mathbf{r}}{\partial z} = \hat{\boldsymbol{k}}
\end{aligned}$$

Aside on notation: some texts use the notation $\hat{\boldsymbol{r}}, \hat{\boldsymbol{\phi}}, \ldots$ to represent the unit vectors that form the local basis set. Though I prefer the notation used here, where the basis vectors are written as $\hat{\boldsymbol{e}}$ with appropriate subscripts (as used in Riley *et al*), you should be aware of, and comfortable with, either possibility.

form a basis set in which we may describe infinitessimal vector displacements in the position of P, $d\mathbf{r}$. It is more usual, however, first to normalise the vectors to obtain their corresponding unit vectors, $\hat{\mathbf{e}}_r$, $\hat{\mathbf{e}}_\phi$, $\hat{\mathbf{e}}_z$. Following the usual rules of calculus we may write:

$$\begin{aligned}
d\mathbf{r} &= \frac{\partial \mathbf{r}}{\partial r}dr + \frac{\partial \mathbf{r}}{\partial \phi}d\phi + \frac{\partial \mathbf{r}}{\partial z}dz \\
&= dr\mathbf{e}_r + d\phi\mathbf{e}_\phi + dz\mathbf{e}_z \\
&= dr\hat{\mathbf{e}}_r + rd\phi\hat{\mathbf{e}}_\phi + dz\hat{\mathbf{e}}_z
\end{aligned}$$

Now here is the important thing to note. In cartesian coordinates, a small change

in (eg) x while keeping y and z constant would result in a displacement of

$$ds = |d\mathbf{r}| = \sqrt{d\mathbf{r}.d\mathbf{r}} = \sqrt{dx^2 + 0 + 0} = dx$$

But in cylindrical polars, a small change in ϕ of $d\phi$ while keeping r and z constant results in a displacement of

$$ds = |d\mathbf{r}| = \sqrt{r^2(d\phi)^2} = r\,d\phi$$

Thus the size of the (infinitessimal) displacement is dependent on the value of r. Factors such as this r are known as **scale factors** or **metric coefficients**, and we must be careful to take them into account when, eg, performing line, surface or volume integrals, as you will below. For cylindrical polars the metric coefficients are clearly $1, r$ and 1.

Example: line integral in cylindrical coordinates

Q Evaluate $\oint_C \mathbf{a} \cdot d\mathbf{l}$, where $\mathbf{a} = x^3\hat{\boldsymbol{j}} - y^3\hat{\boldsymbol{i}} + x^2y\hat{\boldsymbol{k}}$ and C is the circle of radius r in the $z = 0$ plane, centred on the origin.

A Consider figure 4.5. In this case our cylindrical coordinates effectively reduce to plane polars since the path of integration is a circle in the $z = 0$ plane, but let's persist with the full set of coordinates anyway; the $\hat{\boldsymbol{k}}$ component of $\mathbf{a}$ will play no role (it is normal to the path of integration and therefore cancels as seen below).

On the circle of interest

$$\mathbf{a} = r^3(-\sin^3\phi\hat{\boldsymbol{i}} + \cos^3\phi\hat{\boldsymbol{j}} + \cos^2\phi\sin\phi\hat{\boldsymbol{k}})$$

and (since dz = dr = 0 on the path)

$$\begin{aligned} d\mathbf{r} &= r\ d\phi\ \hat{\mathbf{e}}_\phi \\ &= r\,d\phi(-\sin\phi\hat{\boldsymbol{i}} + \cos\phi\hat{\boldsymbol{j}}) \end{aligned}$$

so that

$$\oint_C \mathbf{a} \cdot d\mathbf{r} = \int_0^{2\pi} r^4(\sin^4\phi + \cos^4\phi)d\phi = \frac{3\pi}{2}r^4$$

since

$$\int_0^{2\pi} \sin^4\phi d\phi = \int_0^{2\pi} \cos^4\phi d\phi = \frac{3\pi}{4}$$

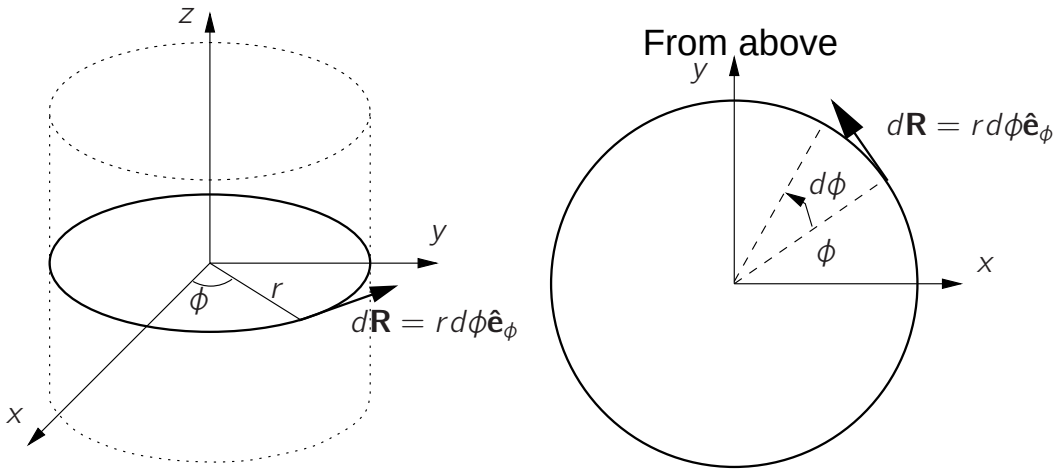


Figure 4.5: Line integral example in cylindrical coordinates

Volume integrals in cylindrical polars

In Cartesian coordinates a volume element is given by (see figure 4.6a):

$$dV = dxdydz$$

Recall that the volume of a parallelopiped is given by the scalar triple product of the vectors which define it (see section 2.1.2). Thus the formula above can be derived (even though it is "obvious") as:

$$dV = dx\hat{\boldsymbol{\imath}}.(dy\hat{\boldsymbol{\jmath}} \times dz\hat{\boldsymbol{k}}) = dxdydz$$

since the basis set is orthonormal.

In cylindrical polars a volume element is given by (see figure 4.6b):

$$dV = dr\hat{\mathbf{e}}_r.(rd\phi\hat{\mathbf{e}}_\phi \times dz\hat{\mathbf{e}}_z) = rd\phi drdz$$

Note also that this volume, because it is a scalar triple product, can be written as a determinant:

$$dV = \begin{vmatrix} \hat{\mathbf{e}}_r dr \\ \hat{\mathbf{e}}_\phi rd\phi \\ \hat{\mathbf{e}}_z dz \end{vmatrix} = \begin{vmatrix} \mathbf{e}_r dr \\ \mathbf{e}_\phi d\phi \\ \mathbf{e}_z dz \end{vmatrix} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial \phi} & \frac{\partial y}{\partial \phi} & \frac{\partial z}{\partial \phi} \\ \frac{\partial x}{\partial z} & \frac{\partial y}{\partial z} & \frac{\partial z}{\partial z} \end{vmatrix} drd\phi dz$$

where the equality on the right-hand side follows from the definitions of $\hat{\mathbf{e}}_r = \frac{\partial \mathbf{r}}{\partial r} = \frac{\partial x}{\partial r}\hat{\boldsymbol{\imath}} + \frac{\partial y}{\partial r}\hat{\boldsymbol{\jmath}} + \frac{\partial z}{\partial r}\hat{\boldsymbol{k}}$, etc. This is the explanation for the "magical" appearance of the determinant in change-of-variables integration that you encountered in your first year maths!

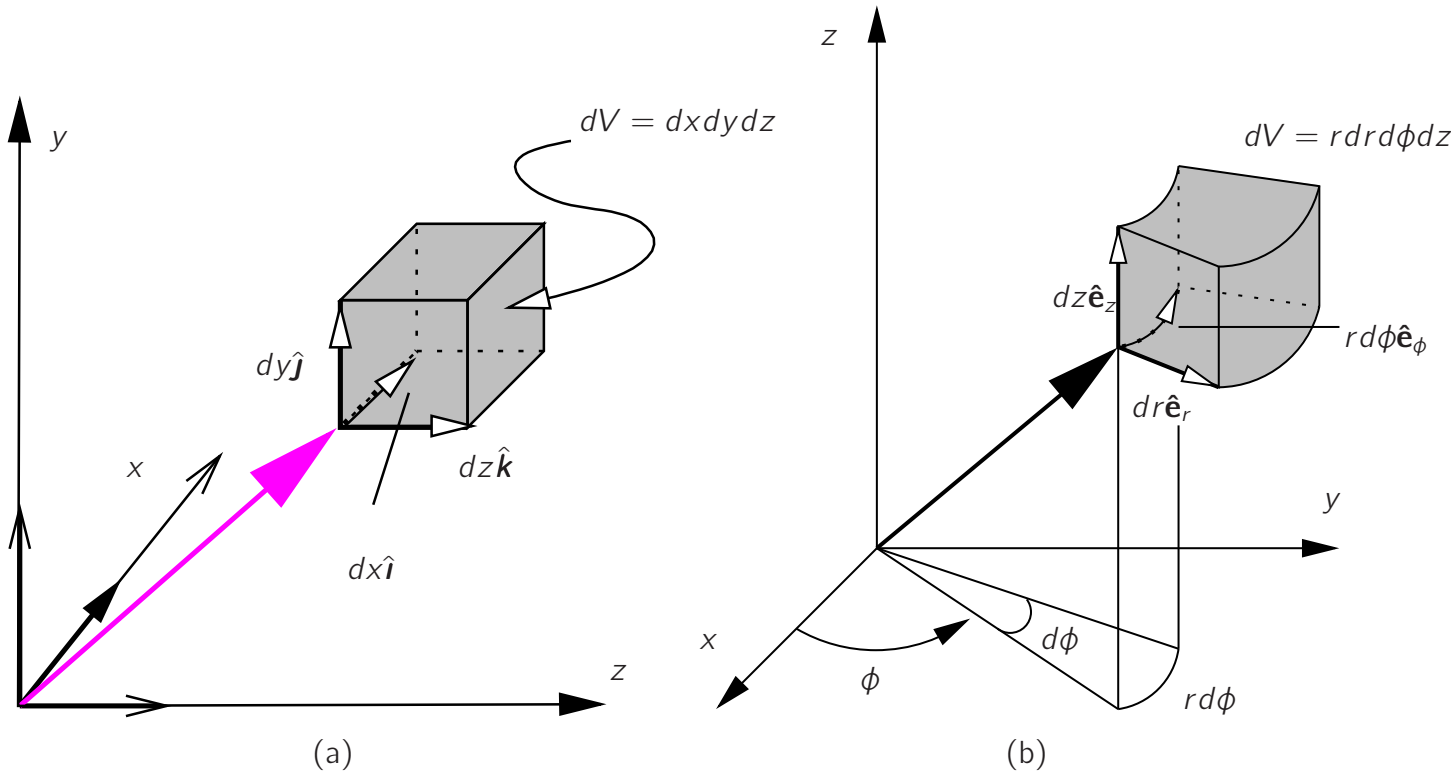


Figure 4.6: Volume elements dV in (a) Cartesian coordinates; (b) Cylindrical polar coordinates

Surface integrals in cylindrical polars

Recall from section 4.4 that for a surface element with normal along $\hat{\boldsymbol{\imath}}$ we have:

$$d\mathbf{S} = dydz\hat{\boldsymbol{\imath}}$$

More explicitly this comes from finding normal to the plane that is tangent to the surface of constant x and from finding the area of an infinitessimal area element on the plane. In this case the plane is spanned by the vectors $\hat{\boldsymbol{\jmath}}$ and $\hat{\boldsymbol{k}}$ and the area of the element given by (see section 1.3):

$$dS = \left| dy\hat{\boldsymbol{\jmath}} \times dz\hat{\boldsymbol{k}} \right|$$

Thus

$$d\mathbf{S} = dy\hat{\boldsymbol{\jmath}} \times dz\hat{\boldsymbol{k}} = \hat{\boldsymbol{\imath}} dS = dydz\hat{\boldsymbol{\imath}}$$

In cylindrical polars, surface area elements (see figure 4.7) are given by:

$$\begin{aligned} d\mathbf{S} &= dr\hat{\mathbf{e}}_r \times rd\phi\hat{\mathbf{e}}_\phi = rdrd\phi\hat{\mathbf{e}}_z \qquad \text{(for surfaces of constant } z\text{)} \\ d\mathbf{S} &= rd\phi\hat{\mathbf{e}}_\phi \times dz\hat{\mathbf{e}}_z = rd\phi dz\hat{\mathbf{e}}_r \qquad \text{(for surfaces of constant } r\text{)} \end{aligned}$$

Similarly we can find $d\mathbf{S}$ for surfaces of constant ϕ, though since these aren't as common this is left as a (relatively easy) exercise.

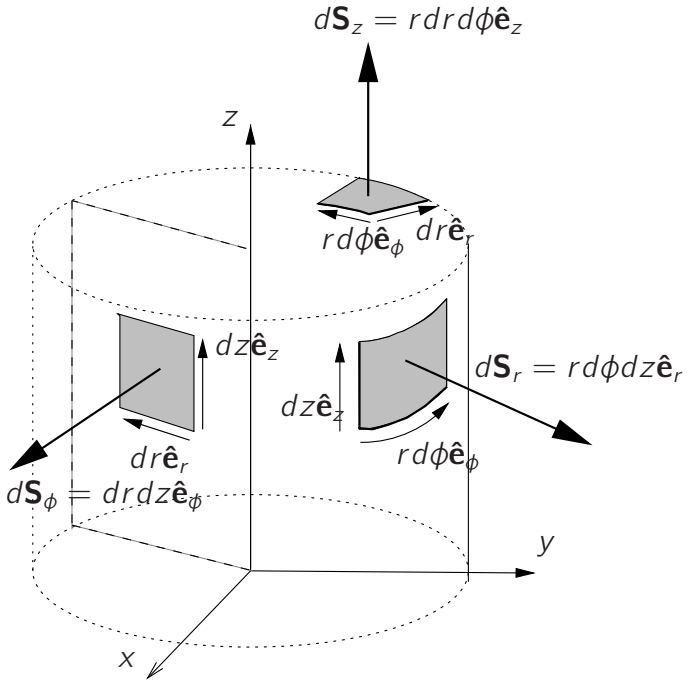


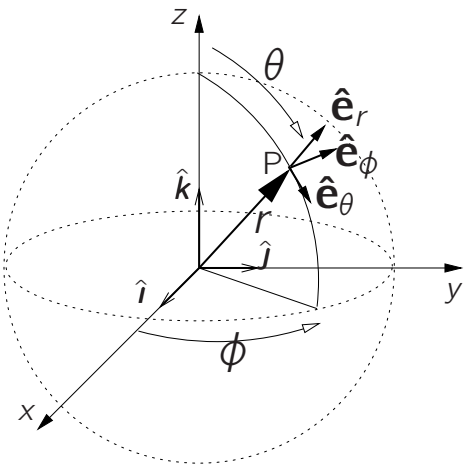
Figure 4.7: Surface elements in cylindrical polar coordinates

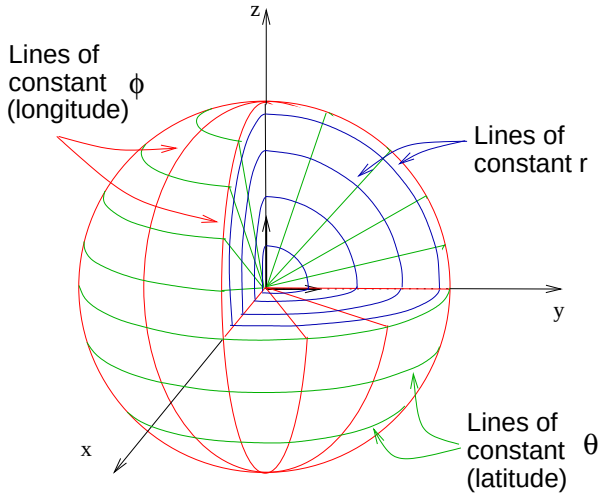
4.6.2 Spherical polars

Much of the development for spherical polars is similar to that for cylindrical polars. As shown in figure 4.6.2 a point in space P having cartesian coordinates x, y, z can be expressed in terms of spherical polar coordinates, r, θ, ϕ as follows:

$$\begin{aligned}\mathbf{r} &= x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}} \\ &= r\sin\theta\cos\phi\hat{\boldsymbol{\imath}} + r\sin\theta\sin\phi\hat{\boldsymbol{\jmath}} + r\cos\theta\hat{\boldsymbol{k}}\end{aligned}$$

The basis set in spherical polars is obtained in an analogous fashion: we find unit





vectors which are in the direction of increase of each coordinate:

$$\begin{aligned}
\mathbf{e}_r = \frac{\partial \mathbf{r}}{\partial r} &= \sin\theta\cos\phi\hat{\boldsymbol{i}} + \sin\theta\sin\phi\hat{\boldsymbol{j}} + \cos\theta\hat{\boldsymbol{k}} &&= \hat{\mathbf{e}}_r \\
\mathbf{e}_\theta = \frac{\partial \mathbf{r}}{\partial \theta} &= r\cos\theta\cos\phi\hat{\boldsymbol{i}} + r\cos\theta\sin\phi\hat{\boldsymbol{j}} - r\sin\theta\hat{\boldsymbol{k}} &&= r\hat{\mathbf{e}}_\theta \\
\mathbf{e}_\phi = \frac{\partial \mathbf{r}}{\partial \phi} &= -r\sin\theta\sin\phi\hat{\boldsymbol{i}} + r\sin\theta\cos\phi\hat{\boldsymbol{j}} &&= r\sin\theta\hat{\mathbf{e}}_\phi
\end{aligned}$$

As with cylindrical polars, it is easily verified that the vectors $\hat{\mathbf{e}}_r, \hat{\mathbf{e}}_\theta, \hat{\mathbf{e}}_\phi$ form an orthonormal basis.

A small displacement $d\mathbf{r}$ is given by:

$$\begin{aligned}
d\mathbf{r} &= \frac{\partial \mathbf{r}}{\partial r}dr + \frac{\partial \mathbf{r}}{\partial \theta}d\theta + \frac{\partial \mathbf{r}}{\partial \phi}d\phi \\
&= dr\mathbf{e}_r + d\theta\mathbf{e}_\theta + d\phi\mathbf{e}_\phi \\
&= dr\hat{\mathbf{e}}_r + rd\theta\hat{\mathbf{e}}_\theta + r\sin\theta d\phi\hat{\mathbf{e}}_\phi
\end{aligned}$$

Thus the metric coefficients are $1, r, r\sin\theta$.

Volume integrals in spherical polars

In spherical polars a volume element is given by (see figure 4.8):

$$dV = dr\hat{\mathbf{e}}_r.(rd\theta\hat{\mathbf{e}}_\theta \times r\sin\theta d\phi\hat{\mathbf{e}}_\phi) = r^2\sin\theta dr d\theta d\phi$$

Note again that this volume could be written as a determinant, but this is left as an exercise.

Surface integrals in spherical polars

The most (the only?) useful surface elements in spherical polars are those tangent to surfaces of constant r (see figure 4.9). The surface direction (unnormalised) is given by $\hat{\mathbf{e}}_\theta \times \hat{\mathbf{e}}_\phi = \hat{\mathbf{e}}_r$ and the area of an infinitessimal surface element is given by $|rd\theta\hat{\mathbf{e}}_\theta \times r\sin\theta d\phi\hat{\mathbf{e}}_\phi| = r^2\sin\theta d\theta d\phi$.

Thus a surface element $d\mathbf{S}$ in spherical polars is given by

$$d\mathbf{S} = rd\theta\hat{\mathbf{e}}_\theta \times r\sin\theta d\phi\hat{\mathbf{e}}_\phi = r^2\sin\theta\hat{\mathbf{e}}_r$$

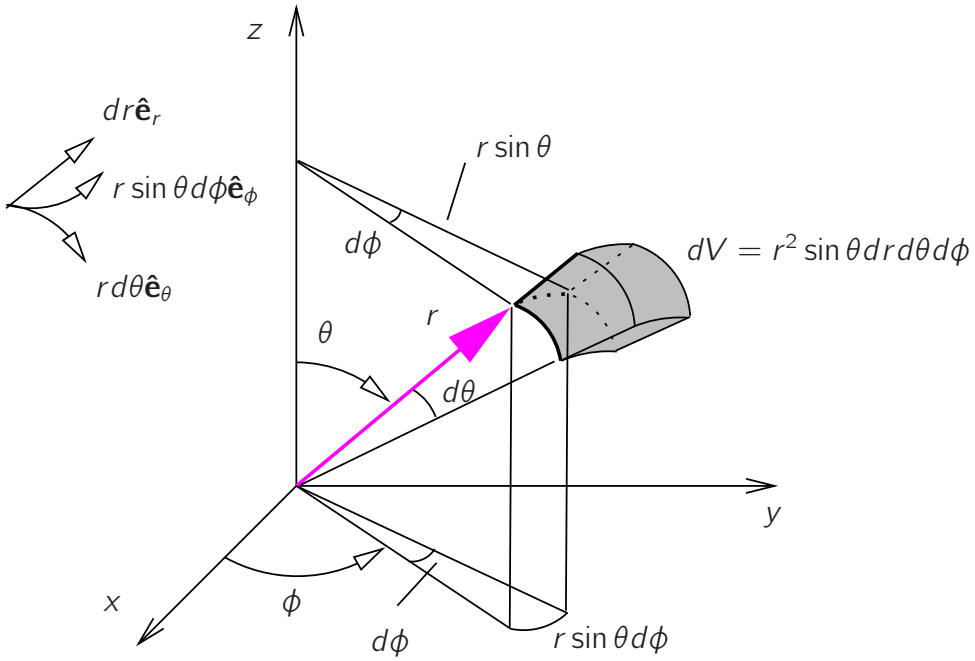


Figure 4.8: Volume element dV in spherical polar coordinates

♣ **Example: surface integral in spherical polars**

Q Evaluate $\int_S \mathbf{a}\cdot d\mathbf{S}$, where $\mathbf{a} = z^3\hat{\boldsymbol{k}}$ and S is the sphere of radius A centred on the origin.

A On the surface of the sphere:

$$\mathbf{a} = A^3 cos^3\theta\hat{\boldsymbol{k}} \qquad d\mathbf{S} = A^2\sin\theta\, d\theta\, d\phi\hat{\mathbf{e}}_r$$

Hence

$$\begin{aligned}\int_S \mathbf{a}\cdot d\mathbf{S} &= \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} A^3 cos^3\theta\; A^2\sin\theta\; [\hat{\mathbf{e}}_r\cdot\hat{\boldsymbol{k}}]\; d\theta d\phi \\ &= A^5\int_0^{2\pi} d\phi\;\int_0^{\pi} cos^3\theta\sin\theta[\cos\theta]\; d\theta \\ &= 2\pi A^5\frac{1}{5}\left[-\cos^5\theta\right]_0^{\pi} \\ &= \frac{4\pi A^5}{5}\end{aligned}$$

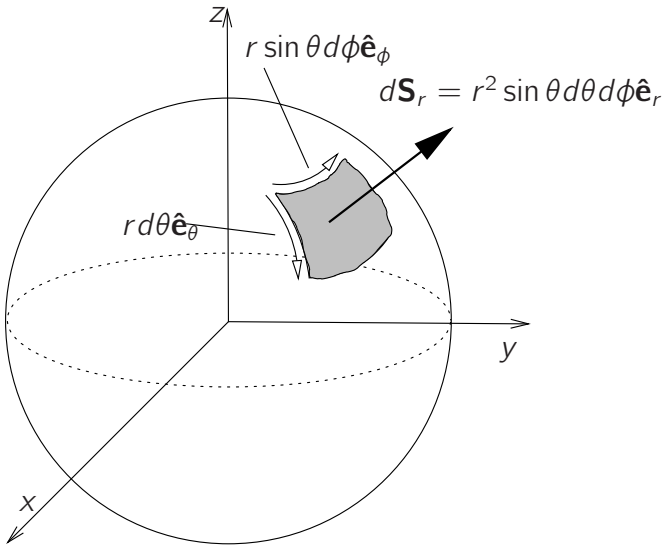


Figure 4.9: Surface element $d\mathbf{S}$ in spherical polar coordinates

4.6.3 General curvilinear coordinates

Cylindrical and spherical polar coordinates are two (useful) examples of general curvilinear coordinates. In general a point P with Cartesian coordinates x, y, z can be expressed in terms of the curvilinear coordinates u, v, w where

$$x = x(u, v, w), \quad y = y(u, v, w), \quad z = z(u, v, w)$$

Thus

$$\mathbf{r} = x(u, v, w)\hat{\boldsymbol{\imath}} + y(u, v, w)\hat{\boldsymbol{\jmath}} + z(u, v, w)\hat{\boldsymbol{k}}$$

and

$$\frac{\partial \mathbf{r}}{\partial u} = \frac{\partial x}{\partial u}\hat{\boldsymbol{\imath}} + \frac{\partial y}{\partial u}\hat{\boldsymbol{\jmath}} + \frac{\partial z}{\partial u}\hat{\boldsymbol{k}}$$

and similarly for partials with respect to v and w, so

$$d\mathbf{r} = \frac{\partial \mathbf{r}}{\partial u}du + \frac{\partial \mathbf{r}}{\partial v}dv + \frac{\partial \mathbf{r}}{\partial w}dw$$

We now define the local coordinate system as before by considering the directions in which each coordinate "unilaterally" (and instantaneously) increases:

$$\begin{aligned}
\mathbf{e}_u &= \frac{\partial \mathbf{r}}{\partial u} = \left|\frac{\partial \mathbf{r}}{\partial u}\right| \hat{\mathbf{e}}_u = h_u \hat{\mathbf{e}}_u \\
\mathbf{e}_v &= \frac{\partial \mathbf{r}}{\partial v} = \left|\frac{\partial \mathbf{r}}{\partial v}\right| \hat{\mathbf{e}}_v = h_v \hat{\mathbf{e}}_v \\
\mathbf{e}_w &= \frac{\partial \mathbf{r}}{\partial w} = \left|\frac{\partial \mathbf{r}}{\partial w}\right| \hat{\mathbf{e}}_w = h_w \hat{\mathbf{e}}_w
\end{aligned}$$

The metric coefficients are therefore $h_u = |\frac{\partial \mathbf{r}}{\partial u}|$, $h_v = |\frac{\partial \mathbf{r}}{\partial v}|$ and $h_w = |\frac{\partial \mathbf{r}}{\partial w}|$.
A volume element is in general given by

$$dV = h_u du\hat{\mathbf{e}}_u.(h_v dv\hat{\mathbf{e}}_v \times h_w dw\hat{\mathbf{e}}_w)$$

and simplifies *if the coordinate system is orthonormal* (since $\hat{\mathbf{e}}_u.(\hat{\mathbf{e}}_v \times \hat{\mathbf{e}}_w) = 1$) to

$$dV = h_u h_v h_w du dv dw$$

A surface element (normal to constant w, say) is in general

$$d\mathbf{S} = h_u du\hat{\mathbf{e}}_u \times h_v dv\hat{\mathbf{e}}_v$$

and simplifies *if the coordinate system is orthogonal* to

$$d\mathbf{S} = h_u h_v du dv\hat{\mathbf{e}}_w$$

4.6.4 Summary

To summarise:

General curvilinear coordinates

$$\begin{aligned}
x &= x(u,v,w), \qquad y = y(u,v,w), \qquad z = z(u,v,w) \\
\mathbf{r} &= x(u,v,w)\hat{\boldsymbol{\imath}} + y(u,v,w)\hat{\boldsymbol{\jmath}} + z(u,v,w)\hat{\boldsymbol{k}} \\
h_u &= \left|\frac{\partial \mathbf{r}}{\partial u}\right|, \qquad h_v = \left|\frac{\partial \mathbf{r}}{\partial v}\right|, \qquad h_w = \left|\frac{\partial \mathbf{r}}{\partial w}\right| \\
\hat{\mathbf{u}} &= \hat{\mathbf{e}}_u = \frac{1}{h_u}\frac{\partial \mathbf{r}}{\partial u}, \qquad \hat{\mathbf{v}} = \hat{\mathbf{e}}_v = \frac{1}{h_v}\frac{\partial \mathbf{r}}{\partial v}, \qquad \hat{\mathbf{w}} = \hat{\mathbf{e}}_w = \frac{1}{h_w}\frac{\partial \mathbf{r}}{\partial w} \\
d\mathbf{r} &= h_u du\hat{\mathbf{u}} + h_v dv\hat{\mathbf{v}} + h_w dw\hat{\mathbf{w}} \\
dV &= h_u h_v h_w du dv dw\ \hat{\mathbf{u}}.(\hat{\mathbf{v}} \times \hat{\mathbf{w}}) \\
d\mathbf{S} &= h_u h_v du dv\ \hat{\mathbf{u}} \times \hat{\mathbf{v}} \quad \text{(for surface element tangent to constant } w\text{)}
\end{aligned}$$

Plane polar coordinates

$$\begin{aligned}
x &= r\cos\theta, \qquad y = r\sin\theta \\
\mathbf{r} &= r\cos\theta\hat{\boldsymbol{\imath}} + r\sin\theta\hat{\boldsymbol{\jmath}} \\
h_r &= 1, \qquad h_\theta = r \\
\hat{\mathbf{e}}_r &= \cos\theta\hat{\boldsymbol{\imath}} + \sin\theta\hat{\boldsymbol{\jmath}}, \qquad \hat{\mathbf{e}}_\theta = -\sin\theta\hat{\boldsymbol{\imath}} + \cos\theta\hat{\boldsymbol{\jmath}} \\
d\mathbf{r} &= dr\hat{\mathbf{e}}_r + rd\theta\hat{\mathbf{e}}_\theta \\
d\mathbf{S} &= rdrd\theta\hat{\boldsymbol{k}}
\end{aligned}$$

Cylindrical polar coordinates

$$
\begin{aligned}
x &= r\cos\phi, \qquad y = r\sin\phi, \qquad z = z \\
\mathbf{r} &= r\cos\phi\hat{\boldsymbol{\imath}} + r\sin\phi\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}} \\
h_r &= 1, \qquad h_\phi = r, \qquad h_z = 1 \\
\hat{\mathbf{e}}_r &= \cos\phi\hat{\boldsymbol{\imath}} + \sin\phi\hat{\boldsymbol{\jmath}}, \qquad \hat{\mathbf{e}}_\phi = -\sin\phi\hat{\boldsymbol{\imath}} + \cos\phi\hat{\boldsymbol{\jmath}}, \qquad \hat{\mathbf{e}}_z = \hat{\boldsymbol{k}} \\
d\mathbf{r} &= dr\hat{\mathbf{e}}_r + rd\phi\hat{\mathbf{e}}_\phi + dz\hat{\mathbf{e}}_z \\
d\mathbf{S} &= rdrd\phi\hat{\boldsymbol{k}} \quad \text{(on the flat ends)} \\
d\mathbf{S} &= rd\phi dz\hat{\mathbf{e}}_r \quad \text{(on the curved sides)} \\
dV &= rdrd\phi dz
\end{aligned}
$$

Spherical polar coordinates

$$
\begin{aligned}
x &= r\sin\theta\cos\phi, \qquad y = r\sin\theta\sin\phi, \qquad z = r\cos\theta \\
\mathbf{r} &= r\sin\theta\cos\phi\hat{\boldsymbol{\imath}} + r\sin\theta\sin\phi\hat{\boldsymbol{\jmath}} + r\cos\theta\hat{\boldsymbol{k}} \\
h_r &= 1, \qquad h_\theta = r, \qquad h_\phi = r\sin\theta \\
\hat{\mathbf{e}}_r &= \sin\theta\cos\phi\hat{\boldsymbol{\imath}} + \sin\theta\sin\phi\hat{\boldsymbol{\jmath}} + \cos\theta\hat{\boldsymbol{k}} \\
\hat{\mathbf{e}}_\theta &= \cos\theta\cos\phi\hat{\boldsymbol{\imath}} + \cos\theta\sin\phi\hat{\boldsymbol{\jmath}} + \sin\theta\hat{\boldsymbol{k}} \\
\hat{\mathbf{e}}_\phi &= -\sin\phi\hat{\boldsymbol{\imath}} + \cos\phi\hat{\boldsymbol{\jmath}} \\
d\mathbf{r} &= dr\hat{\mathbf{e}}_r + rd\theta\hat{\mathbf{e}}_\theta + r\sin\theta d\phi\hat{\mathbf{e}}_\phi \\
d\mathbf{S} &= r^2\sin\theta drd\theta d\phi\hat{\mathbf{e}}_r \quad \text{(on a spherical surface)} \\
dV &= r^2\sin\theta drd\theta d\phi
\end{aligned}
$$

Lecture 5

Vector Operators: Grad, Div and Curl

In the first lecture of the second part of this course we move more to consider properties of fields. We introduce three field operators which reveal interesting collective field properties, viz.

- the **gradient** of a scalar field,
- the **divergence** of a vector field, and
- the **curl** of a vector field.

There are two points to get over about each:

- The mechanics of taking the grad, div or curl, for which you will need to brush up your multivariate calculus.
- The underlying physical meaning — that is, why they are worth bothering about.

In Lecture 6 we will look at combining these vector operators.

5.1 The gradient of a scalar field

Recall the discussion of temperature distribution throughout a room in the overview, where we wondered how a scalar would vary as we moved off in an arbitrary direction. Here we find out how.

If $U(x, y, z)$ is a scalar field, ie a scalar function of position $\mathbf{r} = [x, y, z]$ in 3 dimensions, then its **gradient** at any point is defined in Cartesian co-ordinates by

$$\text{grad}U = \frac{\partial U}{\partial x}\hat{\boldsymbol{\imath}} + \frac{\partial U}{\partial y}\hat{\boldsymbol{\jmath}} + \frac{\partial U}{\partial z}\hat{\boldsymbol{k}} .$$

It is usual to define the **vector operator** which is called "del" or "nabla"

$$\boldsymbol{\nabla} = \hat{\boldsymbol{\imath}}\,\frac{\partial}{\partial x} + \hat{\boldsymbol{\jmath}}\,\frac{\partial}{\partial y} + \hat{\boldsymbol{k}}\,\frac{\partial}{\partial z} .$$

Then

$$\text{grad}U \equiv \boldsymbol{\nabla} U \ .$$

Note immediately that $\boldsymbol{\nabla} U$ is a vector field!

Without thinking too carefully about it, we can see that the gradient of a scalar field tends to point in the direction of greatest change of the field. Later we will be more precise.

♣ Worked examples of gradient evaluation

1. $U = x^2$

$$\Rightarrow \boldsymbol{\nabla} U = \left(\frac{\partial}{\partial x}\hat{\boldsymbol{\imath}} \ + \ \frac{\partial}{\partial y}\hat{\boldsymbol{\jmath}} \ + \ \frac{\partial}{\partial z}\hat{\boldsymbol{k}} \right) x^2 = 2x\hat{\boldsymbol{\imath}} \ .$$

---

2. $U = r^2$

$$\begin{aligned} r^2 &= x^2 + y^2 + z^2 \\ \Rightarrow \boldsymbol{\nabla} U &= \left(\frac{\partial}{\partial x}\hat{\boldsymbol{\imath}} \ + \ \frac{\partial}{\partial y}\hat{\boldsymbol{\jmath}} \ + \ \frac{\partial}{\partial z}\hat{\boldsymbol{k}} \right) (x^2 + y^2 + z^2) \\ &= 2x\hat{\boldsymbol{\imath}} + 2y\hat{\boldsymbol{\jmath}} + 2z\hat{\boldsymbol{k}} \ = \ 2\,\mathbf{r} \ . \end{aligned}$$

---

3. $U = \mathbf{c} \cdot \mathbf{r}$, where $\mathbf{c}$ is constant.

$$\Rightarrow \boldsymbol{\nabla} U = \left(\hat{\boldsymbol{\imath}} \, \frac{\partial}{\partial x} \ + \ \hat{\boldsymbol{\jmath}} \, \frac{\partial}{\partial y} \ + \ \hat{\boldsymbol{k}} \, \frac{\partial}{\partial z} \right) (c_1 x + c_2 y + c_3 z) = c_1\hat{\boldsymbol{\imath}} + c_2\hat{\boldsymbol{\jmath}} + c_3\hat{\boldsymbol{k}} = \mathbf{c} \ .$$

---

4. $U = f(r)$, where $r = \sqrt{(x^2 + y^2 + z^2)}$

U is a function of r alone so df/dr exists. As $U = f(x, y, z)$ also,

$$\frac{\partial f}{\partial x} = \frac{df}{dr}\frac{\partial r}{\partial x} \qquad \frac{\partial f}{\partial y} = \frac{df}{dr}\frac{\partial r}{\partial y} \qquad \frac{\partial f}{\partial z} = \frac{df}{dr}\frac{\partial r}{\partial z} \ .$$

$$\Rightarrow \boldsymbol{\nabla} U = \frac{\partial f}{\partial x}\hat{\boldsymbol{\imath}} + \frac{\partial f}{\partial y}\hat{\boldsymbol{\jmath}} + \frac{\partial f}{\partial z}\hat{\boldsymbol{k}} \quad = \frac{df}{dr}\left(\frac{\partial r}{\partial x}\hat{\boldsymbol{\imath}} + \frac{\partial r}{\partial y}\hat{\boldsymbol{\jmath}} + \frac{\partial r}{\partial z}\hat{\boldsymbol{k}} \right)$$

But $r = \sqrt{x^2 + y^2 + z^2}$, so $\partial r/\partial x = x/r$ and similarly for y, z.

$$\Rightarrow \boldsymbol{\nabla} U = \frac{df}{dr}\left(\frac{x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}}{r} \right) = \frac{df}{dr}\left(\frac{\mathbf{r}}{r} \right) \ .$$

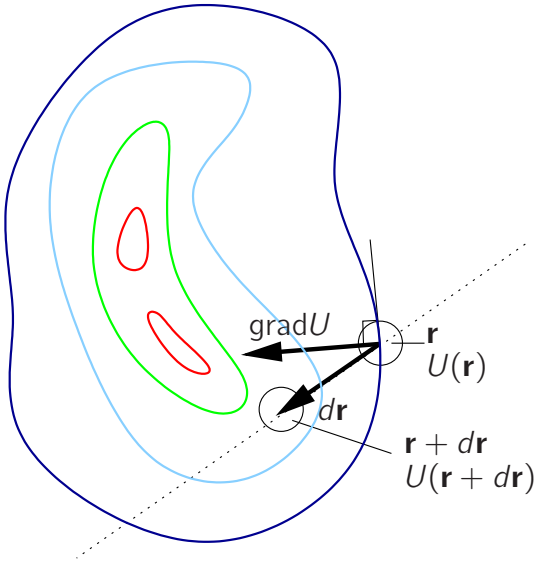


Figure 5.1: The directional derivative

5.2 The significance of grad

If our current position is $\mathbf{r}$ in some scalar field U (Fig. 5.1), and we move an infinitesimal distance $d\mathbf{r}$, we know that the change in U is

$$dU = \frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy + \frac{\partial U}{\partial z}dz \ .$$

But we know that $d\mathbf{r} = (\hat{\boldsymbol{\imath}}dx + \hat{\boldsymbol{\jmath}}dy + \hat{\boldsymbol{k}}dz)$ and $\boldsymbol{\nabla}U = (\hat{\boldsymbol{\imath}}\partial U/\partial x + \hat{\boldsymbol{\jmath}}\partial U/\partial y + \hat{\boldsymbol{k}}\partial U/\partial z)$, so that the change in U is also given by the scalar product

$$dU = \boldsymbol{\nabla}U \cdot d\mathbf{r} \ .$$

Now divide both sides by ds

$$\frac{dU}{ds} = \boldsymbol{\nabla}U \cdot \frac{d\mathbf{r}}{ds} \ .$$

But remember that $|d\mathbf{r}| = ds$, so $d\mathbf{r}/ds$ is a unit vector in the direction of $d\mathbf{r}$.

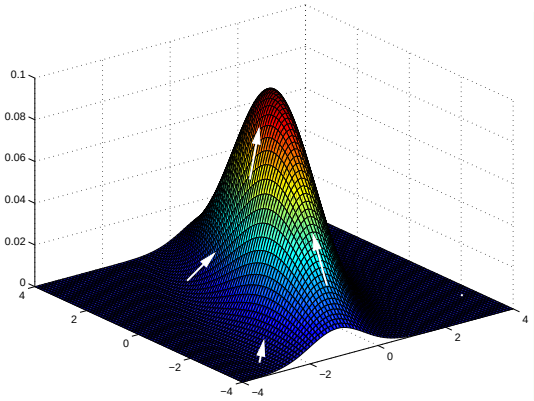
This result can be paraphrased as:

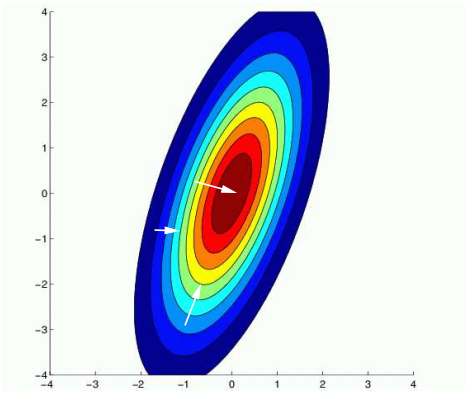
- grad$U$ has the property that the rate of change of U wrt distance in a particular direction ($\hat{\mathbf{d}}$) is the projection of grad$U$ onto that direction (or the component of grad$U$ in that direction).

The quantity dU/ds is called a **directional derivative**, but note that in general it has a different value for each direction, and so has no meaning until you specify the direction.

We could also say that

- At any point P, grad$U$ points in the direction of greatest change of U at P, and has magnitude equal to the rate of change of U wrt distance in that direction.





Another nice property emerges if we think of a surface of constant U – that is the locus (x, y, z) for

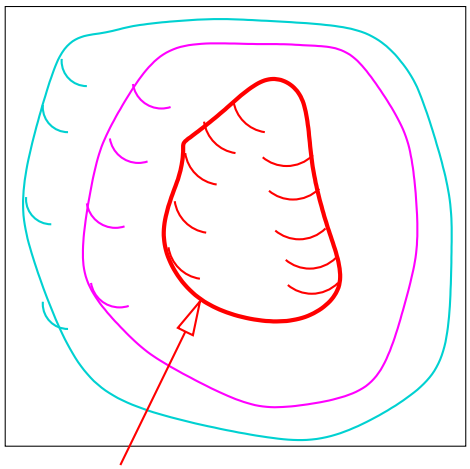
$$U(x, y, z) = \text{constant} \ .$$

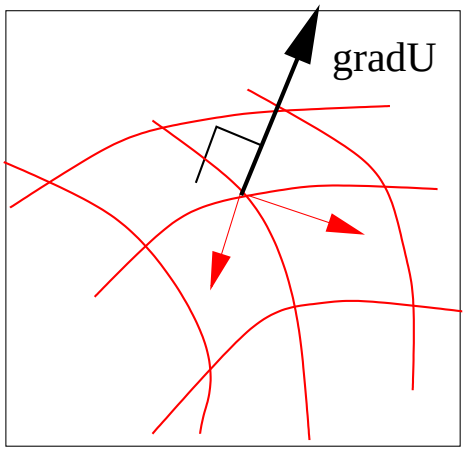
If we move a tiny amount within that iso-U surface, there is no change in U, so $dU/ds = 0$. So for any $d\mathbf{r}/ds$ in the surface

$$\boldsymbol{\nabla} U \cdot \frac{d\mathbf{r}}{ds} \ = \ 0 \ .$$

But $d\mathbf{r}/ds$ is a tangent to the surface, so this result shows that

- grad$U$ is everywhere NORMAL to a surface of constant U.





5.3 The divergence of a vector field

The divergence computes a scalar quantity from a vector field by differentiation.

If $\mathbf{a}(x, y, z)$ is a vector function of position in 3 dimensions, that is $\mathbf{a} = a_1\hat{\boldsymbol{\imath}} + a_2\hat{\boldsymbol{\jmath}} + a_3\hat{\boldsymbol{k}}$, then its divergence at any point is defined in Cartesian co-ordinates by

$$\text{div}\mathbf{a} = \frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z}$$

We can write this in a simplified notation using a scalar product with the $\boldsymbol{\nabla}$ vector differential operator:

$$\text{div}\mathbf{a} = \left(\hat{\boldsymbol{\imath}}\,\frac{\partial}{\partial x} \;+\; \hat{\boldsymbol{\jmath}}\,\frac{\partial}{\partial y} \;+\; \hat{\boldsymbol{k}}\,\frac{\partial}{\partial z}\right)\cdot\mathbf{a} = \boldsymbol{\nabla}\cdot\mathbf{a}$$

Notice that the divergence of a vector field is a scalar field.

♣ Examples of divergence evaluation

	a	div**a**
1)	$x\hat{\boldsymbol{\imath}}$	1
2)	$\mathbf{r}(= x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}})$	3
3)	$\mathbf{r}/r^3$	0
4)	$r\mathbf{c}$, for $\mathbf{c}$ constant	$(\mathbf{r}\cdot\mathbf{c})/r$

We work through example 3).

The x component of $\mathbf{r}/r^3$ is $x.(x^2+y^2+z^2)^{-3/2}$, and we need to find $\partial/\partial x$ of it.

$$\begin{aligned}\frac{\partial}{\partial x}x.(x^2+y^2+z^2)^{-3/2} &= 1.(x^2+y^2+z^2)^{-3/2} + x\frac{-3}{2}(x^2+y^2+z^2)^{-5/2}.2x \\ &= r^{-3}\left(1-3x^2r^{-2}\right)\ .\end{aligned}$$

The terms in y and z are similar, so that

$$\begin{aligned}\text{div}(\mathbf{r}/r^3) &= r^{-3}\left(3-3(x^2+y^2+z^2)r^{-2}\right) = r^{-3}\left(3-3\right) \\ &= 0\end{aligned}$$

5.4 The significance of div

Consider a typical vector field, water flow, and denote it by $\mathbf{a}(\mathbf{r})$. This vector has magnitude equal to the mass of water crossing a unit area perpendicular to the direction of $\mathbf{a}$ per unit time.

Now take an infinitesimal volume element dV and figure out the balance of the flow of $\mathbf{a}$ in and out of dV.

To be specific, consider the volume element $dV = dxdydz$ in Cartesian coordinates, and think first about the face of area $dxdz$ perpendicular to the y axis and facing outwards in the negative y direction. (That is, the one with surface area $d\mathbf{S} = -dxdz\hat{\boldsymbol{j}}$.)

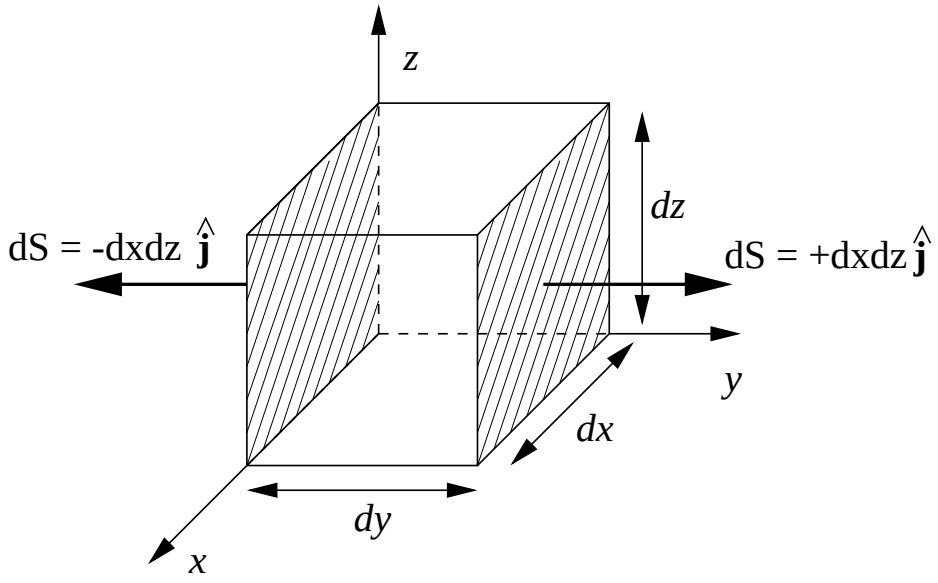


Figure 5.2: Elemental volume for calculating divergence.

The component of the vector $\mathbf{a}$ normal to this face is $\mathbf{a} \cdot \hat{\boldsymbol{j}} = a_y$, and is pointing inwards, and so its contribution to the OUTWARD flux from this surface is

$$\mathbf{a} \cdot d\mathbf{S} \quad = \quad - a_y(y) dz dx \ ,$$

where $a_y(y)$ means that a_y is a function of y. (By the way, flux here denotes mass per unit time.)

A similar contribution, but of opposite sign, will arise from the opposite face, but we must remember that we have moved along y by an amount dy, so that this OUTWARD amount is

$$a_y(y + dy) dz dx = \left(a_y + \frac{\partial a_y}{\partial y} dy \right) dx dz$$

The total outward amount from these two faces is

$$\frac{\partial a_y}{\partial y} dy dx dz = \frac{\partial a_y}{\partial y} dV$$

Summing the other faces gives a total outward flux of

$$\left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \right) dV = \boldsymbol{\nabla} \cdot \mathbf{a} \ \ dV$$

So we see that

The divergence of a vector field represents the flux generation per unit volume at each point of the field. (Divergence because it is an efflux not an influx.)

Interestingly we also saw that the total efflux from the infinitesimal volume was equal to the flux integrated over the surface of the volume.

(NB: The above does not constitute a rigorous proof of the assertion because we have not proved that the quantity calculated is independent of the co-ordinate system used, but it will suffice for our purposes.)

5.5 The Laplacian: div(grad$U$) of a scalar field

Recall that grad$U$ of *any* scalar field U is a vector field. Recall also that we can compute the divergence of any vector field. So we can certainly compute div(grad$U$), even if we don't know what it means yet.

Here is where the $\boldsymbol{\nabla}$ operator starts to be really handy.

$$\begin{aligned}
\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}U) &= \left(\hat{\boldsymbol{\imath}}\,\frac{\partial}{\partial x} + \hat{\boldsymbol{\jmath}}\,\frac{\partial}{\partial y} + \hat{\boldsymbol{k}}\,\frac{\partial}{\partial z}\right)\cdot\left(\left(\hat{\boldsymbol{\imath}}\,\frac{\partial}{\partial x} + \hat{\boldsymbol{\jmath}}\,\frac{\partial}{\partial y} + \hat{\boldsymbol{k}}\,\frac{\partial}{\partial z}\right)U\right)\\
&= \left(\left(\hat{\boldsymbol{\imath}}\,\frac{\partial}{\partial x} + \hat{\boldsymbol{\jmath}}\,\frac{\partial}{\partial y} + \hat{\boldsymbol{k}}\,\frac{\partial}{\partial z}\right)\cdot\left(\hat{\boldsymbol{\imath}}\,\frac{\partial}{\partial x} + \hat{\boldsymbol{\jmath}}\,\frac{\partial}{\partial y} + \hat{\boldsymbol{k}}\,\frac{\partial}{\partial z}\right)\right)U\\
&= \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)U\\
&= \left(\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}+\frac{\partial^2 U}{\partial z^2}\right)
\end{aligned}$$

This last expression occurs frequently in engineering science (you will meet it next in solving Laplace's Equation in partial differential equations). For this reason, the operator ∇^2 is called the "Laplacian"

$$\nabla^2 U = \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)U$$

Laplace's equation itself is

$$\nabla^2 U = 0$$

♣ **Examples of $\nabla^2 U$ evaluation**

	U	$\nabla^2 U$
1)	$r^2 (= x^2 + y^2 + z^2)$	6
2)	xy^2z^3	$2xz^3 + 6xy^2z$
3)	$1/r$	0

Let's prove example (3) (which is particularly significant – can you guess why?).

$$1/r = (x^2 + y^2 + z^2)^{-1/2}$$

$$\begin{aligned} \frac{\partial}{\partial x}\frac{\partial}{\partial x}(x^2 + y^2 + z^2)^{-1/2} &= \frac{\partial}{\partial x} - x.(x^2 + y^2 + z^2)^{-3/2} \\ &= -(x^2 + y^2 + z^2)^{-3/2} + 3x.x.(x^2 + y^2 + z^2)^{-5/2} \\ &= (1/r^3)(-1 + 3x^2/r^2) \end{aligned}$$

Adding up similar terms for y and z

$$\nabla^2 \frac{1}{r} = \frac{1}{r^3}\left(-3 + 3\frac{(x^2 + y^2 + x^2)}{r^2}\right) = 0$$

5.6 The curl of a vector field

So far we have seen the operator $\boldsymbol{\nabla}$ applied to a scalar field $\boldsymbol{\nabla} U$; and dotted with a vector field $\boldsymbol{\nabla} \cdot \mathbf{a}$.

We are now overwhelmed by an irrestible temptation to

- cross it with a vector field $\boldsymbol{\nabla} \times \mathbf{a}$

This gives the **curl of a vector field**

$$\boldsymbol{\nabla} \times \mathbf{a} \equiv \text{curl}(\mathbf{a})$$

We can follow the pseudo-determinant recipe for vector products, so that

$$\begin{aligned} \boldsymbol{\nabla} \times \mathbf{a} &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_x & a_y & a_z \end{vmatrix} \quad \text{(remember it this way)} \\ &= \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right)\hat{\boldsymbol{\imath}} + \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right)\hat{\boldsymbol{\jmath}} + \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right)\hat{\boldsymbol{k}} \end{aligned}$$

♣ Examples of curl evaluation

	a	$\nabla \times \mathbf{a}$
1)	$-y\hat{\boldsymbol{i}} + x\hat{\boldsymbol{j}}$	$2\hat{\boldsymbol{k}}$
2)	$x^2y^2\hat{\boldsymbol{k}}$	$2x^2y\hat{\boldsymbol{i}} - 2xy^2\hat{\boldsymbol{j}}$

5.7 The significance of curl

Perhaps the first example gives a clue. The field $\mathbf{a} = -y\hat{\boldsymbol{i}} + x\hat{\boldsymbol{j}}$ is sketched in Figure 5.3(a). (It is the field you would calculate as the velocity field of an object rotating with $\boldsymbol{\omega} = [0, 0, 1]$.) This field has a curl of $2\hat{\boldsymbol{k}}$, which is in the r-h screw sense out of the page. You can also see that a field like this must give a finite value to the line integral around the complete loop $\oint_C \mathbf{a} \cdot d\mathbf{r}$.

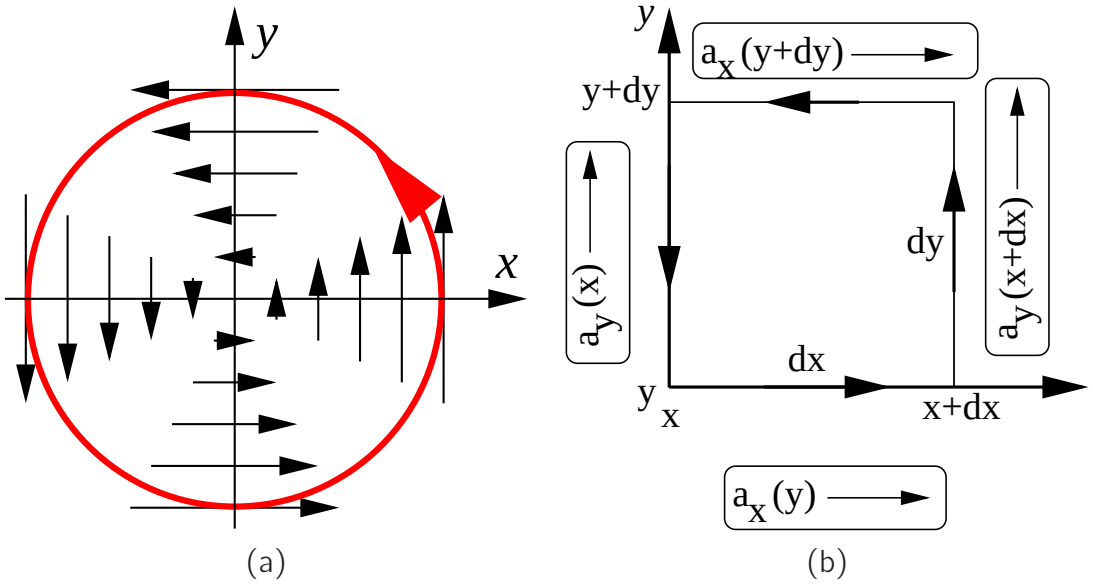


Figure 5.3: (a) A rough sketch of the vector field $-y\hat{\boldsymbol{i}} + x\hat{\boldsymbol{j}}$. (b) An element in which to calculate curl.

In fact curl is closely related to the line integral around a loop.

The **circulation** of a vector **a** round any closed curve C is defined to be $\oint_C \mathbf{a} \cdot d\mathbf{r}$
and the **curl** of the vector field **a** represents the **vorticity**, or **circulation per unit area**, of the field.

Our proof uses the small rectangular element dx by dy shown in Figure 5.3(b). Consider the circulation round the perimeter of a rectangular element.

The fields in the x direction at the bottom and top are

$$a_x(y) \qquad \text{and} \qquad a_x(y+dy) = a_x(y) + \frac{\partial a_x}{\partial y} dy,$$

where $a_x(y)$ denotes a_x is a function of y, and the fields in the y direction at the left and right are

$$a_y(x) \qquad \text{and} \qquad a_y(x+dx) = a_y(x) + \frac{\partial a_y}{\partial x} dx$$

Starting at the bottom and working round in the anticlockwise sense, the four contributions to the circulation dC are therefore as follows, where the minus signs take account of the path being opposed to the field:

$$\begin{aligned} dC &= +[a_x(y)\ dx] + [a_y(x+dx)\ dy] - [a_x(y+dy)\ dx] - [a_y(x)\ dy] \\ &= +[a_x(y)\ dx] + \left[\left(a_y(x) + \frac{\partial a_y}{\partial x} dx\right) dy\right] - \left[\left(a_x(y) + \frac{\partial a_x}{\partial y} dy\right) dx\right] - [a_y(x)\ dy] \\ &= \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right)\ dx\ dy \\ &= (\boldsymbol{\nabla} \times \mathbf{a}) \cdot d\mathbf{S} \end{aligned}$$

where $d\mathbf{S} = dxdy\hat{\boldsymbol{k}}$.

NB: Again, this is not a completely rigorous proof as we have not shown that the result is independent of the co-ordinate system used.

5.8 Some definitions involving div, curl and grad

- A vector field with zero divergence is said to be **solenoidal**.
- A vector field with zero curl is said to be **irrotational**.
- A scalar field with zero gradient is said to be, er, **constant**.

Lecture 6

Vector Operator Identities

In this lecture we look at more complicated identities involving vector operators. The main thing to appreciate it that the operators behave both as vectors and as differential operators, so that the usual rules of taking the derivative of, say, a product must be observed.

There could be a cottage industry inventing vector identities. HLT contains a lot of them. So why not leave it at that?

First, since grad, div and curl describe key aspects of vectors fields, they arise often in practice, and so the identities can save you a lot of time and hacking of partial derivatives, as we will see when we consider Maxwell's equation as an example later.

Secondly, they help to identify other practically important vector operators. So, although this material is a bit dry, the relevance of the identities should become clear later in other Engineering courses.

6.1 Identity 1: curl grad $U = \mathbf{0}$

$$\begin{aligned}
\boldsymbol{\nabla} \times \boldsymbol{\nabla} U &= \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ \partial U/\partial x & \partial U/\partial y & \partial U/\partial z \end{vmatrix} \\
&= \hat{\boldsymbol{\imath}} \left(\frac{\partial^2 U}{\partial y \partial z} - \frac{\partial^2 U}{\partial z \partial y} \right) + \hat{\boldsymbol{\jmath}} \left(\right) + \hat{\boldsymbol{k}} \left(\right) \\
&= \mathbf{0} \ ,
\end{aligned}$$

as $\partial^2/\partial y \partial z = \partial^2/\partial z \partial y$.

Note that the output is a null *vector*.

6.2 Identity 2: div curl $\mathbf{a} = 0$

$$\begin{aligned}
\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\times\mathbf{a} &= \begin{vmatrix} \partial/\partial x & \partial/\partial y & \partial/\partial z \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ a_x & a_y & a_z \end{vmatrix} \\
&= \frac{\partial^2 a_z}{\partial x\partial y} - \frac{\partial^2 a_y}{\partial x\partial z} - \frac{\partial^2 a_z}{\partial y\partial x} + \frac{\partial^2 a_x}{\partial y\partial z} + \frac{\partial^2 a_y}{\partial z\partial x} - \frac{\partial^2 a_x}{\partial z\partial y} \\
&= 0
\end{aligned}$$

6.3 Identity 3: div and curl of $U\mathbf{a}$

Suppose that $U(\mathbf{r})$ is a scalar field and that $\mathbf{a}(\mathbf{r})$ is a vector field and we are interested in the product $U\mathbf{a}$. This is a vector field, so we can compute its divergence and curl. For example the density $\rho(\mathbf{r})$ of a fluid is a scalar field, and the instantaneous velocity of the fluid $\mathbf{v}(\mathbf{r})$ is a vector field, and we are probably interested in mass flow rates for which we will be interested in $\rho(\mathbf{r})\mathbf{v}(\mathbf{r})$.

The divergence (a scalar) of the product $U\mathbf{a}$ is given by:

$$\begin{aligned}
\boldsymbol{\nabla}\cdot(U\mathbf{a}) &= U(\boldsymbol{\nabla}\cdot\mathbf{a}) + (\boldsymbol{\nabla}U)\cdot\mathbf{a} \\
&= U\text{div}\mathbf{a} + (\text{grad}U)\cdot\mathbf{a}
\end{aligned}$$

In a similar way, we can take the curl of the vector field $U\mathbf{a}$, and the result should be a vector field:

$$\boldsymbol{\nabla}\times(U\mathbf{a}) = U\boldsymbol{\nabla}\times\mathbf{a} + (\boldsymbol{\nabla}U)\times\mathbf{a} \quad .$$

6.4 Identity 4: div of $\mathbf{a}\times\mathbf{b}$

Life quickly gets trickier when vector or scalar products are involved: For example, it is not *that* obvious that

$$\text{div}(\mathbf{a}\times\mathbf{b}) = \text{curl}\mathbf{a}\cdot\mathbf{b} - \mathbf{a}\cdot\text{curl}\mathbf{b}$$

To show this, use the determinant:

$$\begin{aligned}
\begin{vmatrix} \partial/\partial x_i & \partial/\partial x_j & \partial/\partial x_k \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} &= \frac{\partial}{\partial x}[a_y b_z - a_z b_y] + \frac{\partial}{\partial y}[a_z b_x - a_x b_z] + \frac{\partial}{\partial z}[a_x b_y - a_y b_x] \\
&= \ldots \text{ bash out the products } \ldots \\
&= \text{curl}\mathbf{a}\cdot\mathbf{b} - \mathbf{a}\cdot(\text{curl }\mathbf{b})
\end{aligned}$$

6.5 Identity 5: curl($\mathbf{a} \times \mathbf{b}$)

$$\text{curl}(\mathbf{a} \times \mathbf{b}) = \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ a_y b_z - a_z b_y & a_z b_x - a_x b_z & a_x b_y - a_y b_x \end{vmatrix}$$

so the $\hat{\boldsymbol{\imath}}$ component is

$$\frac{\partial}{\partial y}(a_x b_y - a_y b_x) - \frac{\partial}{\partial z}(a_z b_x - a_x b_z)$$

which can be written as the sum of four terms:

$$a_x\left(\frac{\partial b_y}{\partial y} + \frac{\partial b_z}{\partial z}\right) - b_x\left(\frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right) + \left(b_y\frac{\partial}{\partial y} + b_z\frac{\partial}{\partial z}\right)a_x - \left(a_y\frac{\partial}{\partial y} + a_z\frac{\partial}{\partial z}\right)b_x$$

Adding $a_x(\partial b_x/\partial x)$ to the first of these, and subtracting it from the last, and doing the same with $b_x(\partial a_x/\partial x)$ to the other two terms, we find that (you should of course check this):

$$\boldsymbol{\nabla} \times (\mathbf{a} \times \mathbf{b}) = (\boldsymbol{\nabla} \cdot \mathbf{b})\mathbf{a} - (\boldsymbol{\nabla} \cdot \mathbf{a})\mathbf{b} + [\mathbf{b} \cdot \boldsymbol{\nabla}]\mathbf{a} - [\mathbf{a} \cdot \boldsymbol{\nabla}]\mathbf{b}$$

where $[\mathbf{a} \cdot \boldsymbol{\nabla}]$ can be regarded as new, and very useful, scalar differential operator.

6.6 Definition of the operator $[\mathbf{a} \cdot \boldsymbol{\nabla}]$

This is a *scalar operator*, but it can obviously can be applied to a scalar field, resulting in a scalar field, or to a vector field resulting in a vector field:

$$[\mathbf{a} \cdot \boldsymbol{\nabla}] \equiv \left[a_x\frac{\partial}{\partial x} + a_y\frac{\partial}{\partial y} + a_z\frac{\partial}{\partial z}\right] \quad .$$

6.7 Identity 6: curl(curl$\mathbf{a}$) for you to derive

The following important identity is stated, and left as an exercise:

$$\text{curl}(\text{curl}\mathbf{a}) = \text{graddiv}\mathbf{a} - \nabla^2\mathbf{a}$$

where

$$\nabla^2\mathbf{a} = \nabla^2 a_x\hat{\boldsymbol{\imath}} + \nabla^2 a_y\hat{\boldsymbol{\jmath}} + \nabla^2 a_z\hat{\boldsymbol{k}}$$

♣ Example of Identity 6: electromagnetic waves

Q: James Clerk Maxwell established a set of four vector equations which are fundamental to working out how electromagnetic waves propagate. The entire telecommunications industry is built on these.

$$\begin{aligned} \text{div}\mathbf{D} &= \rho \\ \text{div}\mathbf{B} &= 0 \\ \text{curl}\mathbf{E} &= -\frac{\partial}{\partial t}\mathbf{B} \\ \text{curl}\mathbf{H} &= \mathbf{J} + \frac{\partial}{\partial t}\mathbf{D} \end{aligned}$$

In addition, we can assume the following, which should all be familiar to you:
$\mathbf{B} = \mu_r\mu_0\mathbf{H}, \qquad \mathbf{J} = \sigma\mathbf{E}, \qquad \mathbf{D} = \epsilon_r\epsilon_0\mathbf{E},$
where all the scalars are constants.

Now show that in a material with zero free charge density, $\rho = 0$, and with zero conductivity, $\sigma = 0$, the electric field $\mathbf{E}$ must be a solution of the wave equation

$$\nabla^2\mathbf{E} = \mu_r\mu_0\epsilon_r\epsilon_0(\partial^2\mathbf{E}/\partial t^2) \ .$$

A: First, a bit of respect. Imagine you are the first to do this — this is a tingle moment.

$$\begin{aligned} \text{div}\mathbf{D} &= \text{div}(\epsilon_r\epsilon_0\mathbf{E}) = \epsilon_r\epsilon_0\text{div}\mathbf{E} = \rho = 0 \quad \Rightarrow \text{div}\mathbf{E} = 0. \quad (a) \\ \text{div}\mathbf{B} &= \text{div}(\mu_r\mu_0\mathbf{H}) = \mu_r\mu_0\text{div}\mathbf{H} = 0 \quad \Rightarrow \text{div}\mathbf{B} = 0 \quad (b) \\ \text{curl}\mathbf{E} &= -\partial\mathbf{B}/\partial t = -\mu_r\mu_0(\partial\mathbf{H}/\partial t) \quad (c) \\ \text{curl}\mathbf{H} &= \mathbf{J} + \partial\mathbf{D}/\partial t = \mathbf{0} + \epsilon_r\epsilon_0(\partial\mathbf{E}/\partial t) \quad (d) \end{aligned}$$

But we know (or rather you worked out in Identity 6) that curlcurl = graddiv$-\nabla^2$, and using (c)

$$\text{curlcurl}\mathbf{E} = \text{graddiv}\mathbf{E} - \nabla^2\mathbf{E} = \text{curl}\left(-\mu_r\mu_0(\partial\mathbf{H}/\partial t)\right)$$

so interchanging the order of partial differentation, and using (a) div$\mathbf{E} = 0$:

$$\begin{aligned} -\nabla^2\mathbf{E} &= -\mu_r\mu_0\frac{\partial}{\partial t}(\text{curl}\mathbf{H}) \\ &= -\mu_r\mu_0\frac{\partial}{\partial t}\left(\epsilon_r\epsilon_0\frac{\partial\mathbf{E}}{\partial t}\right) \\ \Rightarrow \nabla^2\mathbf{E} &= \mu_r\mu_0\epsilon_r\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2} \end{aligned}$$

This equation is actually three equations, one for each component:

$$\nabla^2 E_x = \mu_r \mu_0 \epsilon_r \epsilon_0 \frac{\partial^2 E_x}{\partial t^2}$$

and so on for E_y and E_z.

6.8 Grad, div, curl and ∇^2 in curvilinear co-ordinate systems

It is possible to obtain general expressions for grad, div and curl in any orthogonal curvilinear co-ordinate system by making use of the h factors which were introduced in Lecture 4.

We recall that the unit vector in the direction of increasing u, with v and w being kept constant, is

$$\hat{\mathbf{u}} = \frac{1}{h_u}\frac{\partial \mathbf{r}}{\partial u}$$

where $\mathbf{r}$ is the position vector, and

$$h_u = \left|\frac{\partial \mathbf{r}}{\partial u}\right|$$

is the metric coefficient. Similar expressions apply for the other co-ordinate directions. Then

$$d\mathbf{r} = h_u \hat{\mathbf{u}} du + h_v \hat{\mathbf{v}} dv + h_w \hat{\mathbf{w}} dw \ .$$

6.9 Grad in curvilinear coordinates

Noting that $U = U(\mathbf{r})$ and $U = U(u, v, w)$, and using the properties of the gradient of a scalar field obtained previously

$$\boldsymbol{\nabla} U \cdot d\mathbf{r} = dU = \frac{\partial U}{\partial u} du + \frac{\partial U}{\partial v} dv + \frac{\partial U}{\partial w} dw$$

It follows that

$$\boldsymbol{\nabla} U \cdot (h_u \hat{\mathbf{u}} du + h_v \hat{\mathbf{v}} dv + h_w \hat{\mathbf{w}} dw) = \frac{\partial U}{\partial u} du + \frac{\partial U}{\partial v} dv + \frac{\partial U}{\partial w} dw$$

The only way this can be satisfied for independent du, dv, dw is when

$$\boxed{\boldsymbol{\nabla} U = \frac{1}{h_u}\frac{\partial U}{\partial u}\hat{\mathbf{u}} + \frac{1}{h_v}\frac{\partial U}{\partial v}\hat{\mathbf{v}} + \frac{1}{h_w}\frac{\partial U}{\partial w}\hat{\mathbf{w}}}$$

6.10 Divergence in curvilinear coordinates

Expressions can be obtained for the divergence of a vector field in orthogonal curvilinear co-ordinates by making use of the flux property.

We consider an element of volume dV. If the curvilinear coordinates are orthogonal then the little volume is a cuboid (to first order in small quantities) and

$$dV = h_u\ h_v\ h_w\ \ du\ dv\ dw\ \ .$$

However, it is not quite a cuboid: the area of two opposite faces will differ as the scale parameters are functions of u, v and w in general.

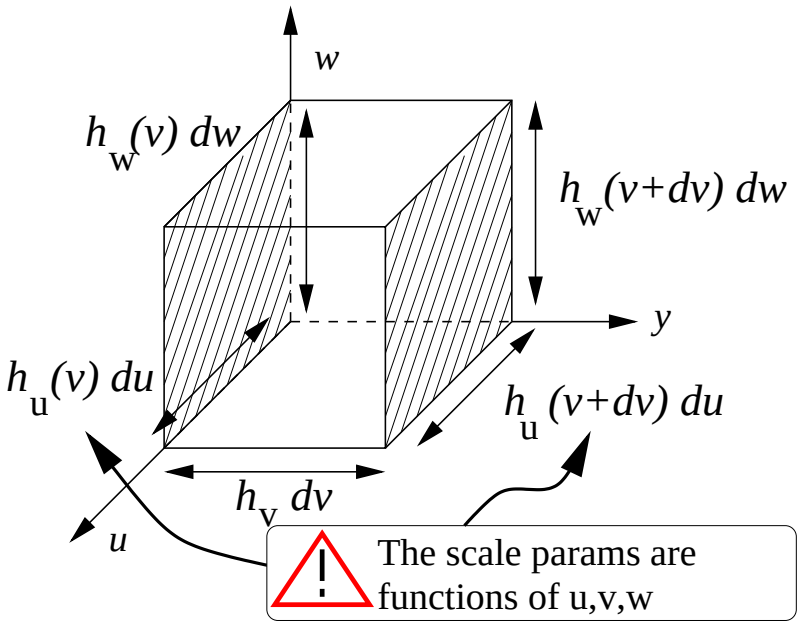


Figure 6.1: Elemental volume for calculating divergence in orthogonal curvilinear coordinates

So the net efflux from the two faces in the $\hat{\mathbf{v}}$ direction shown in Figure 6.1 is

$$\begin{aligned} &= \left[a_v + \frac{\partial a_v}{\partial v}dv\right]\left[h_u + \frac{\partial h_u}{\partial v}dv\right]\left[h_w + \frac{\partial h_w}{\partial v}dv\right] dudw - a_v h_u h_w dudw \\ &= \frac{\partial(a_v h_u h_w)}{\partial v} dudvdw \end{aligned}$$

which is easily shown by multiplying the first line out and dropping second order terms (i.e. $(dv)^2$).

By definition div is the net efflux per unit volume, so summing up the other faces:

$$\text{div}\mathbf{a}\ dV = \left(\frac{\partial(a_u\ h_v\ h_w)}{\partial u} + \frac{\partial(a_v\ h_u\ h_w)}{\partial v} + \frac{\partial(a_w\ h_u\ h_v)}{\partial w}\right) dudvdw$$

$$\Rightarrow \text{div}\mathbf{a}\ h_u h_v h_w\ dudvdw = \left(\frac{\partial(a_u\ h_v\ h_w)}{\partial u} + \frac{\partial(a_v\ h_u\ h_w)}{\partial v} + \frac{\partial(a_w\ h_u\ h_v)}{\partial w}\right) dudvdw$$

So, finally,

$$\text{div}\mathbf{a} = \frac{1}{h_u h_v h_w}\left(\frac{\partial(a_u\ h_v\ h_w)}{\partial u} + \frac{\partial(a_v\ h_u\ h_w)}{\partial v} + \frac{\partial(a_w\ h_u\ h_v)}{\partial w}\right)$$

6.11 Curl in curvilinear coordinates

Recall from Lecture 5 that we computed the z component of curl as the circulation per unit area from

$$dC = \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right)\ dx\ dy$$

By analogy with our derivation of divergence, you will realize that for an orthogonal curvilinear coordinate system we can write the area as $h_u h_v du dw$. But the opposite sides are no longer quite of the same length. The lower of the pair in Figure 6.2 is length $h_u(v)du$, but the upper is of length $h_u(v+dv)du$

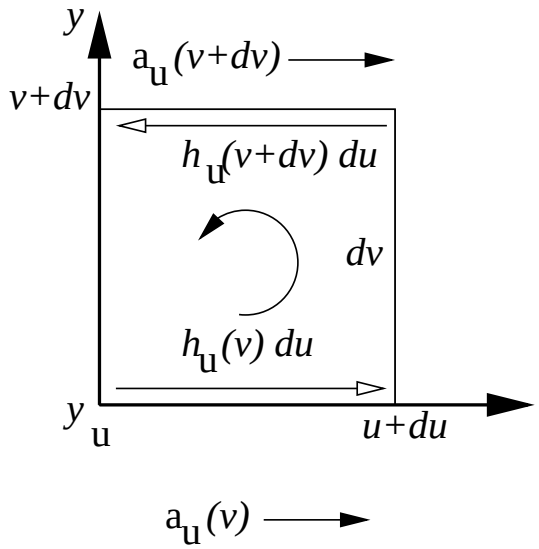


Figure 6.2: Elemental loop for calculating curl in orthogonal curvilinear coordinates

Summing this pair gives a contribution to the circulation

$$a_u(v)h_u(v)du - a_u(v+dv)h_u(v+dv)du = -\frac{\partial(h_u a_u)}{\partial v}dvdu$$

and together with the other pair:

$$dC = \left(-\frac{\partial(h_u a_u)}{\partial v} + \frac{\partial(h_v a_v)}{\partial u}\right)dudv$$

So the circulation per unit area is

$$\frac{dC}{h_u h_v du dv} = \frac{1}{h_u h_v}\left(\frac{\partial(h_v a_v)}{\partial u} - \frac{\partial(h_u a_u)}{\partial v}\right)$$

and hence curl is

$$\begin{aligned}\text{curl}\mathbf{a}(u,v,w) \;=\; & \frac{1}{h_v h_w}\left(\frac{\partial(h_w a_w)}{\partial v} - \frac{\partial(h_v a_v)}{\partial w}\right)\hat{\mathbf{u}} \;+ \\ & \frac{1}{h_w h_u}\left(\frac{\partial(h_u a_u)}{\partial w} - \frac{\partial(h_w a_w)}{\partial u}\right)\hat{\mathbf{v}} \;+ \\ & \frac{1}{h_u h_v}\left(\frac{\partial(h_v a_v)}{\partial u} - \frac{\partial(h_u a_u)}{\partial v}\right)\hat{\mathbf{w}}\end{aligned}$$

You should check that this can be written as

Curl in curvilinear coords:

$$\text{curl}\mathbf{a}(u,v,w) \;=\; \frac{1}{h_u h_v h_w}\begin{vmatrix} h_u\hat{\mathbf{u}} & h_v\hat{\mathbf{v}} & h_w\hat{\mathbf{w}} \\ \frac{\partial}{\partial u} & \frac{\partial}{\partial v} & \frac{\partial}{\partial w} \\ h_u a_u & h_v a_v & h_w a_w \end{vmatrix}$$

6.12 The Laplacian in curvilinear coordinates

Substitution of the components of grad$U$ into the expression for div$\mathbf{a}$ immediately (!*?) gives the following expression for the Laplacian in general orthogonal coordinates:

$$\nabla^2 U = \frac{1}{h_u h_v h_w}\left[\frac{\partial}{\partial u}\left(\frac{h_v h_w}{h_u}\frac{\partial U}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{h_w h_u}{h_v}\frac{\partial U}{\partial v}\right) + \frac{\partial}{\partial w}\left(\frac{h_u h_v}{h_w}\frac{\partial U}{\partial w}\right)\right].$$

6.13 Grad Div, Curl, ∇^2 in cylindrical polars

Here $(u,v,w) \to (r,\phi,z)$. The position vector is $\mathbf{r} = r\cos\phi\hat{\boldsymbol{\imath}} + r\sin\phi\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}$, and $h_r = |\partial\mathbf{r}/\partial r|$, etc.

$$\begin{aligned}\Rightarrow h_r &= \sqrt{(\cos^2\phi + sin^2\phi)} = 1, \\ h_\phi &= \sqrt{(r^2\sin^2\phi + r^2\cos^2\phi)} = r, \\ h_z &= 1\end{aligned}$$

$$
\begin{aligned}
\Rightarrow \text{grad}U &= \frac{\partial U}{\partial r}\hat{\mathbf{e}}_r + \frac{1}{r}\frac{\partial U}{\partial \phi}\hat{\mathbf{e}}_\phi + \frac{\partial U}{\partial z}\hat{\boldsymbol{k}} \\
\text{div}\mathbf{a} &= \frac{1}{r}\left(\frac{\partial (r a_r)}{\partial r} + \frac{\partial a_\phi}{\partial \phi}\right) + \frac{\partial a_z}{\partial z} \\
\text{curl}\mathbf{a} &= \left(\frac{1}{r}\frac{\partial a_z}{\partial \phi} - \frac{\partial a_\phi}{\partial z}\right)\hat{\mathbf{e}}_r + \left(\frac{\partial a_r}{\partial z} - \frac{\partial a_z}{\partial r}\right)\hat{\mathbf{e}}_\phi + \frac{1}{r}\left(\frac{\partial (r a_\phi)}{\partial r} - \frac{\partial a_r}{\partial \phi}\right)\hat{\boldsymbol{k}} \\
\nabla^2 U &= \text{Tutorial Exercise}
\end{aligned}
$$

6.14 Grad Div, Curl, ∇^2 in spherical polars

Here $(u, v, w) \to (r, \theta, \phi)$. The position vector is $\mathbf{r} = r\sin\theta\cos\phi\hat{\boldsymbol{i}} + r\sin\theta\sin\phi\hat{\boldsymbol{j}} + r\cos\theta\hat{\boldsymbol{k}}$.

$$
\begin{aligned}
\Rightarrow h_r &= \sqrt{(\sin^2\theta(\cos^2\phi + \sin^2\phi) + \cos^2\theta)} = 1 \\
h_\theta &= \sqrt{(r^2\cos^2\theta(\cos^2\phi + \sin^2\phi) + r^2\sin^2\theta)} = r \\
h_\phi &= \sqrt{(r^2\sin^2\theta(\sin^2\phi + \cos^2\phi)} = r\sin\theta
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \text{grad}U &= \frac{\partial U}{\partial r}\hat{\mathbf{e}}_r + \frac{1}{r}\frac{\partial U}{\partial \theta}\hat{\mathbf{e}}_\theta + \frac{1}{r\sin\theta}\frac{\partial U}{\partial \phi}\hat{\mathbf{e}}_\phi \\
\text{div}\mathbf{a} &= \frac{1}{r^2}\frac{\partial (r^2 a_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial (a_\theta \sin\theta)}{\partial \theta} + \frac{1}{r\sin\theta}\frac{\partial a_\phi}{\partial \phi} \\
\text{curl}\mathbf{a} &= \frac{\hat{\mathbf{e}}_r}{r\sin\theta}\left(\frac{\partial}{\partial \theta}(a_\phi\sin\theta) - \frac{\partial}{\partial \phi}(a_\theta)\right) + \frac{\hat{\mathbf{e}}_\theta}{r\sin\theta}\left(\frac{\partial}{\partial \phi}(a_r) - \frac{\partial}{\partial r}(a_\phi r\sin\theta)\right) + \\
&\qquad \frac{\hat{\mathbf{e}}_\phi}{r}\left(\frac{\partial}{\partial r}(a_\theta r) - \frac{\partial}{\partial \theta}(a_r)\right) \\
\nabla^2 U &= \text{Tutorial Exercise}
\end{aligned}
$$

♣ Examples

Q1 Find curl$\mathbf{a}$ in (i) Cartesians and (ii) Spherical polars when $\mathbf{a} = x(x\hat{\boldsymbol{i}} + y\hat{\boldsymbol{j}} + z\hat{\boldsymbol{k}})$.

A1 (i) In Cartesians

$$
\text{curl}\mathbf{a} = \begin{vmatrix} \hat{\boldsymbol{i}} & \hat{\boldsymbol{j}} & \hat{\boldsymbol{k}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ x^2 & xy & xz \end{vmatrix} = -z\hat{\boldsymbol{j}} + y\hat{\boldsymbol{k}} \ .
$$

(ii) In spherical polars, $x = r\sin\theta\cos\phi$ and $\mathbf{r} = (x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}})$. So

$$\begin{aligned} \mathbf{a} &= r^2\sin\theta\cos\phi\hat{\mathbf{e}}_r \\ \Rightarrow a_r &= r^2\sin\theta\cos\phi; \qquad a_\theta = 0; \qquad a_\phi = 0\ . \end{aligned}$$

Hence as

$$\text{curl}\mathbf{a} = \frac{\hat{\mathbf{e}}_r}{r\sin\theta}\left(\frac{\partial}{\partial\theta}(a_\phi\sin\theta) - \frac{\partial}{\partial\phi}(a_\theta)\right) + \frac{\hat{\mathbf{e}}_\theta}{r\sin\theta}\left(\frac{\partial}{\partial\phi}(a_r) - \frac{\partial}{\partial r}(a_\phi r\sin\theta)\right) + \frac{\hat{\mathbf{e}}_\phi}{r}\left(\frac{\partial}{\partial r}(a_\theta r) - \frac{\partial}{\partial\theta}(a_r)\right)$$

$$\begin{aligned} \text{curl}\mathbf{a} &= \frac{\hat{\mathbf{e}}_\theta}{r\sin\theta}\left(\frac{\partial}{\partial\phi}(r^2\sin\theta\cos\phi)\right) + \frac{\hat{\mathbf{e}}_\phi}{r}\left(-\frac{\partial}{\partial\theta}(r^2\sin\theta\cos\phi)\right) \\ &= \frac{\hat{\mathbf{e}}_\theta}{r\sin\theta}(-r^2\sin\theta\sin\phi) + \frac{\hat{\mathbf{e}}_\phi}{r}\left(-r^2\cos\theta\cos\phi\right)) \\ &= \hat{\mathbf{e}}_\theta(-r\sin\phi) + \hat{\mathbf{e}}_\phi(-r\cos\theta\cos\phi) \end{aligned}$$

Checking: these two results should be the same, but to check we need expressions for $\hat{\mathbf{e}}_\theta, \hat{\mathbf{e}}_\phi$ in terms of $\hat{\boldsymbol{\imath}}$ etc.

Remember that we can work out the unit vectors $\hat{\mathbf{e}}_r$ and so on in terms of $\hat{\boldsymbol{\imath}}$ etc using

$$\hat{\mathbf{e}}_r = \frac{1}{h_1}\frac{\partial\mathbf{r}}{dr}; \qquad \hat{\mathbf{e}}_\theta = \frac{1}{h_2}\frac{\partial\mathbf{r}}{d\theta}; \qquad \hat{\mathbf{e}}_\phi = \frac{1}{h_3}\frac{\partial\mathbf{r}}{d\phi} \quad \text{where } \mathbf{r} = x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}\ .$$

Grinding through we find

$$\begin{bmatrix} \hat{\mathbf{e}}_r \\ \hat{\mathbf{e}}_\theta \\ \hat{\mathbf{e}}_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\imath}} \\ \hat{\boldsymbol{\jmath}} \\ \hat{\boldsymbol{k}} \end{bmatrix} = [R] \begin{bmatrix} \hat{\boldsymbol{\imath}} \\ \hat{\boldsymbol{\jmath}} \\ \hat{\boldsymbol{k}} \end{bmatrix}$$

Don't be shocked to see a rotation matrix $[R]$: we are after all rotating one right-handed orthogonal coord system into another.

So the result in spherical polars is

$$\begin{aligned} \text{curl}\mathbf{a} &= (\cos\theta\cos\phi\hat{\boldsymbol{\imath}} + \cos\theta\sin\phi\hat{\boldsymbol{\jmath}} - \sin\theta\hat{\boldsymbol{k}})(-r\sin\phi) + (-\sin\phi\hat{\boldsymbol{\imath}} + \cos\phi\hat{\boldsymbol{\jmath}})(-r\cos\theta\cos \\ &= -r\cos\theta\hat{\boldsymbol{\jmath}} + r\sin\theta\sin\phi\hat{\boldsymbol{k}} \\ &= -z\hat{\boldsymbol{\jmath}} + y\hat{\boldsymbol{k}} \end{aligned}$$

which is exactly the result in Cartesians.

Q2 Find the divergence of the vector field $\mathbf{a} = r\mathbf{c}$ where $\mathbf{c}$ is a constant vector (i) using Cartesian coordinates and (ii) using Spherical Polar coordinates.

A2 (i) Using Cartesian coords:

$$\begin{aligned}\text{div}\mathbf{a} &= \frac{\partial}{\partial x}(x^2+y^2+z^2)^{1/2}c_x + \ldots \\ &= x.(x^2+y^2+z^2)^{-1/2}c_x + \ldots \\ &= \frac{1}{r}\mathbf{r}\cdot\mathbf{c} \quad .\end{aligned}$$

(ii) Using Spherical polars

$$\mathbf{a} = a_r\hat{\mathbf{e}}_r + a_\theta\hat{\mathbf{e}}_\theta + a_\phi\hat{\mathbf{e}}_\phi$$

and our first task is to find a_r and so on. We can't do this by inspection, and finding their values requires more work than you might think! Recall

$$\begin{bmatrix}\hat{\mathbf{e}}_r\\ \hat{\mathbf{e}}_\theta\\ \hat{\mathbf{e}}_\phi\end{bmatrix} = \begin{bmatrix}\sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta\\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta\\ -\sin\phi & \cos\phi & 0\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\imath}}\\ \hat{\boldsymbol{\jmath}}\\ \hat{\boldsymbol{k}}\end{bmatrix} = [R]\begin{bmatrix}\hat{\boldsymbol{\imath}}\\ \hat{\boldsymbol{\jmath}}\\ \hat{\boldsymbol{k}}\end{bmatrix}$$

Now the point is the same point in space whatever the coordinate system, so

$$a_r\hat{\mathbf{e}}_r + a_\theta\hat{\mathbf{e}}_\theta + a_\phi\hat{\mathbf{e}}_\phi = a_x\hat{\boldsymbol{\imath}} + a_y\hat{\boldsymbol{\jmath}} + a_z\hat{\boldsymbol{k}}$$

and using the inner product

$$\begin{bmatrix}a_r\\ a_\theta\\ a_\phi\end{bmatrix}^\top\begin{bmatrix}\hat{\mathbf{e}}_r\\ \hat{\mathbf{e}}_\theta\\ \hat{\mathbf{e}}_\phi\end{bmatrix} = \begin{bmatrix}a_x\\ a_y\\ a_z\end{bmatrix}^\top\begin{bmatrix}\hat{\boldsymbol{\imath}}\\ \hat{\boldsymbol{\jmath}}\\ \hat{\boldsymbol{k}}\end{bmatrix}$$

$$\begin{bmatrix}a_r\\ a_\theta\\ a_\phi\end{bmatrix}^\top[R]\begin{bmatrix}\hat{\boldsymbol{\imath}}\\ \hat{\boldsymbol{\jmath}}\\ \hat{\boldsymbol{k}}\end{bmatrix} = \begin{bmatrix}a_x\\ a_y\\ a_z\end{bmatrix}^\top\begin{bmatrix}\hat{\boldsymbol{\imath}}\\ \hat{\boldsymbol{\jmath}}\\ \hat{\boldsymbol{k}}\end{bmatrix}$$

$$\Rightarrow \begin{bmatrix}a_r\\ a_\theta\\ a_\phi\end{bmatrix}^\top[R] = \begin{bmatrix}a_x\\ a_y\\ a_z\end{bmatrix}^\top$$

$$\Rightarrow \begin{bmatrix}a_r\\ a_\theta\\ a_\phi\end{bmatrix}^\top = \begin{bmatrix}a_x\\ a_y\\ a_z\end{bmatrix}^\top[R]^\top$$

$$\Rightarrow \begin{bmatrix}a_r\\ a_\theta\\ a_\phi\end{bmatrix} = [R]\begin{bmatrix}a_x\\ a_y\\ a_z\end{bmatrix}$$

For our particular problem, $a_x = rc_x$, etc, where c_x is a constant, so now we can write down

$$\begin{aligned} a_r &= r(\sin\theta\cos\phi c_x + \sin\theta\sin\phi c_y + \cos\theta c_z) \\ a_\theta &= r(\cos\theta\cos\phi c_x + \cos\theta\sin\phi c_y - \sin\theta c_z) \\ a_\phi &= r(-\sin\phi c_x + \cos\phi c_y) \end{aligned}$$

Now all we need to do is to bash out

$$\mathrm{div}\mathbf{a} = \frac{1}{r^2}\frac{\partial(r^2 a_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(a_\theta\sin\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial a_\phi}{\partial\phi}$$

In glorious detail this is

$$\begin{aligned} \mathrm{div}\mathbf{a} &= 3(\sin\theta\cos\phi c_x + \sin\theta\sin\phi c_y + \cos\theta c_z) + \\ &\quad \frac{1}{\sin\theta}\left(\cos^2\theta - \sin^2\theta)(\cos\phi c_x + \sin\phi c_y) - 2\sin\theta\cos\theta c_z\right) + \\ &\quad \frac{1}{\sin\theta}(-\cos\phi c_x - \sin\phi c_y) \end{aligned}$$

A bit more bashing and you'll find

$$\begin{aligned} \mathrm{div}\mathbf{a} &= \sin\theta\cos\phi c_x + \sin\theta\sin\phi c_y + \cos\theta c_z \\ &= \hat{\mathbf{e}}_r \cdot \mathbf{c} \end{aligned}$$

This is EXACTLY what you worked out before of course.

Take home messages from these examples:

- Just as physical vectors are independent of their coordinate systems, so are differential operators.
- Don't forget about the vector geometry you did in the 1st year. Rotation matrices are useful!
- Spherical polars were NOT a good coordinate system in which to think about this problem. Let the symmetry guide you.

Lecture 7

Gauss' and Stokes' Theorems

This section finally begins to deliver on why we introduced div grad and curl. Two theorems, both of them over two hundred years old, are explained:

- **Gauss' Theorem** enables an integral taken over a volume to be replaced by one taken over the surface bounding that volume, and vice versa. Why would we want to do that? Computational efficiency and/or numerical accuracy!
- **Stokes' Law** enables an integral taken around a closed curve to be replaced by one taken over *any* surface bounded by that curve.

7.1 Gauss' Theorem

Suppose that $\mathbf{a}(\mathbf{r})$ is a vector field and we want to compute the total flux of the field across the surface S that bounds a volume V. That is, we are interested in calculating:

$$\int_S \mathbf{a} \cdot d\mathbf{S}$$

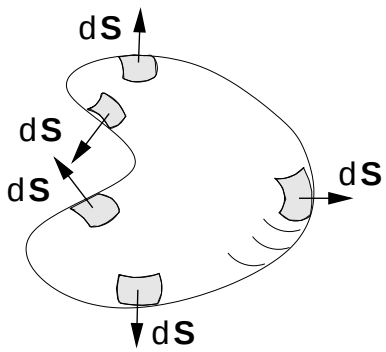


Figure 7.1: The surface element $d\mathbf{S}$ must stick out of the surface.

where recall that $d\mathbf{S}$ is normal to the locally planar surface element and **must everywhere point out of the volume** as shown in Figure 7.1.

Gauss' Theorem tells us that we can do this by considering the total flux generated inside the volume V:

Gauss' Theorem

$$\int_S \mathbf{a} \cdot d\mathbf{S} = \int_V \text{div } \mathbf{a} \; dV$$

obtained by integrating the divergence over the entire volume.

7.2 Informal proof

An non-rigorous proof can be realized by recalling that we defined div by considering the efflux dE from the surfaces of an infinitesimal volume element

$$dE = \mathbf{a} \cdot d\mathbf{S}$$

and defining it as

$$\text{div } \mathbf{a} \; dV \quad = \quad dE \quad = \quad \mathbf{a} \cdot d\mathbf{S} \; .$$

If we sum over the volume elements, this results in a sum over the surface elements. But if two elemental surface touch, their $d\mathbf{S}$ vectors are in opposing direction and cancel as shown in Figure 7.2. Thus the sum over surface elements gives the overall bounding surface.

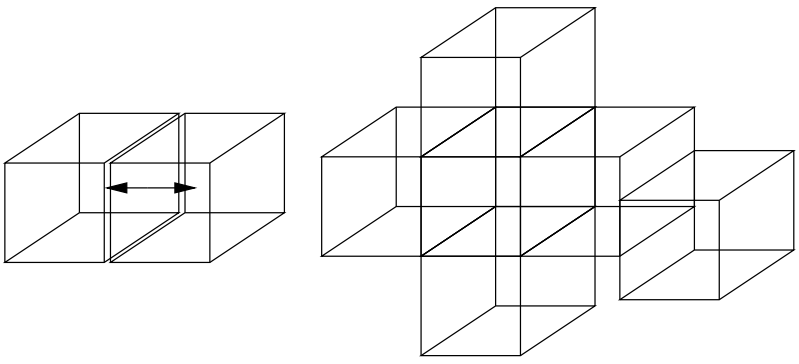


Figure 7.2: When two elements touch, the $d\mathbf{S}$ vectors at the common surface cancel out. One can imagine building the entire volume up from the infinitesimal units.

♣ Example of Gauss' Theorem

This is a typical example, in which the surface integral is rather tedious, whereas the volume integral is straightforward.

Q Derive $\int_S \mathbf{a} \cdot d\mathbf{S}$ where $\mathbf{a} = z^3\hat{\boldsymbol{k}}$ and S is the surface of a sphere of radius R centred on the origin:

1. directly;
2. by applying Gauss' Theorem

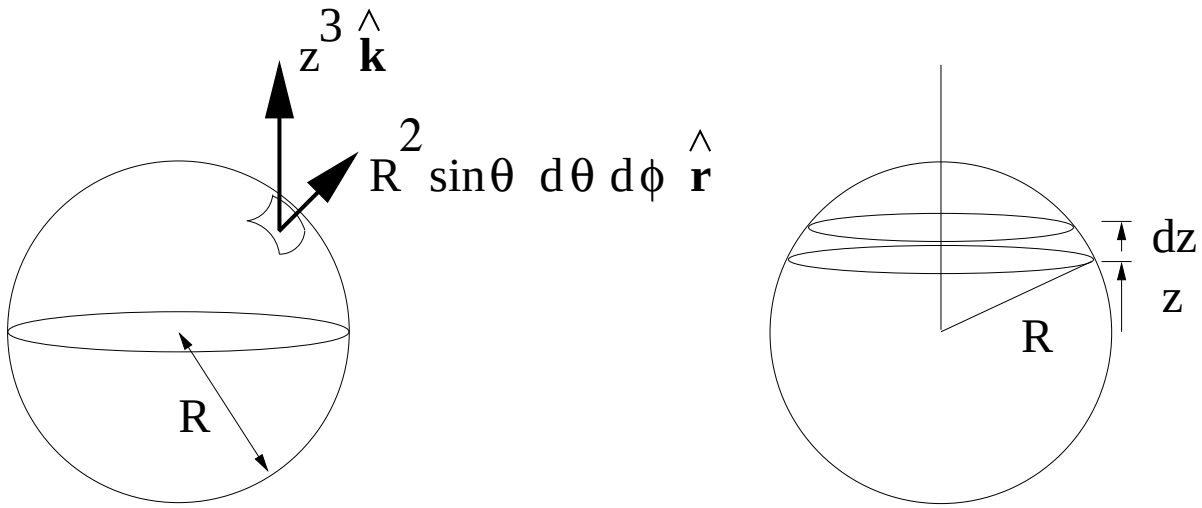


Figure 7.3:

A (1) On the surface of the sphere, $\mathbf{a} = R^3\cos^3\theta\hat{\boldsymbol{k}}$ and $d\mathbf{S} = R^2\sin\theta d\theta d\phi\hat{\mathbf{r}}$. Everywhere $\hat{\mathbf{r}} \cdot \hat{\boldsymbol{k}} = \cos\theta$.

$$\begin{aligned}
\Rightarrow \int_S \mathbf{a} \cdot d\mathbf{S} &= \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} R^3\cos^3\theta \;.\; R^2\sin\theta d\theta d\phi\hat{\mathbf{e}}_r \cdot \hat{\boldsymbol{k}} \\
&= \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} R^3\cos^3\theta \;.\; R^2\sin\theta d\theta d\phi \;.\; \cos\theta \\
&= 2\pi R^5\int_0^{\pi}\cos^4\theta\sin\theta d\theta \\
&= \frac{2\pi R^5}{5}\left[-\cos^5\theta\right]_0^{\pi} = \frac{4\pi R^5}{5}
\end{aligned}$$

(2) To apply Gauss' Theorem, we need to figure out div $\mathbf{a}$ and decide how to compute the volume integral. The first is easy:

$$\text{div}\mathbf{a} = 3z^2$$

For the second, because div$\mathbf{a}$ involves just z, we can divide the sphere into discs of constant z and thickness dz, as shown in Fig. 7.3. Then

$$dV = \pi(R^2 - z^2)dz$$

and

$$\begin{aligned}\int_V \text{div } \mathbf{a} dV &= 3\pi \int_{-R}^{R} z^2(R^2 - z^2)dz \\ &= 3\pi \left[\frac{R^2 z^3}{3} - \frac{z^5}{5}\right]_{-R}^{R} \\ &= \frac{4\pi R^5}{5}\end{aligned}$$

7.3 Surface versus volume integrals

At first sight, it might seem that with a computer performing surface integrals might be better than a volume integral, perhaps because there are, somehow, "fewer elements". However, this is not the case. Imagine doing a surface integral over a wrinkly surface, say that of the moon. All the elements involved in the integration are "difficult" and must be modelled correctly. With a volume integral, most of the elements are not at the surface, and so the bulk of the integral is done without accurate modelling. The computation is easier, faster, and better conditioned numerically.

7.4 Extension to Gauss' Theorem

Suppose the vector field $\mathbf{a}(\mathbf{r})$ is of the form $\mathbf{a} = U(\mathbf{r})\mathbf{c}$, where $U(\mathbf{r})$ as scalar field and $\mathbf{c}$ is a constant vector. Then, as we showed in the previous lecture,

$$\begin{aligned}\text{div } \mathbf{a} &= \text{grad}U \cdot \mathbf{c} + U\text{div } \mathbf{c} \\ &= \text{grad}U \cdot \mathbf{c}\end{aligned}$$

since div$\mathbf{c} = 0$ because $\mathbf{c}$ is constant.

Gauss' Theorem becomes

$$\int_S U\mathbf{c} \cdot d\mathbf{S} = \int_V \text{grad } U \cdot \mathbf{c} dV$$

or, alternatively, taking the constant $\mathbf{c}$ out of the integrals

$$\mathbf{c} \cdot \left(\int_S U d\mathbf{S}\right) = \mathbf{c} \cdot \left(\int_V \text{grad } U dV\right)$$

This is still a scalar equation but we now note that the vector $\mathbf{c}$ is arbitrary so that the result must be true for any vector $\mathbf{c}$. This can be true only if the vector equation

$$\int_S U d\mathbf{S} = \int_V \text{grad } U dV$$

is satisfied.

If you think this is fishy, just write $\mathbf{c} = \hat{\boldsymbol{\imath}}$, then $\mathbf{c} = \hat{\boldsymbol{\jmath}}$, and $\mathbf{c} = \hat{\boldsymbol{k}}$ in turn, and you must obtain the three components of $\int_S U d\mathbf{S}$ in turn.

Further "extensions" can be obtained of course. For example one might be able to write the vector field of interest as

$$\mathbf{a}(\mathbf{r}) = \mathbf{b}(\mathbf{r}) \times \mathbf{c}$$

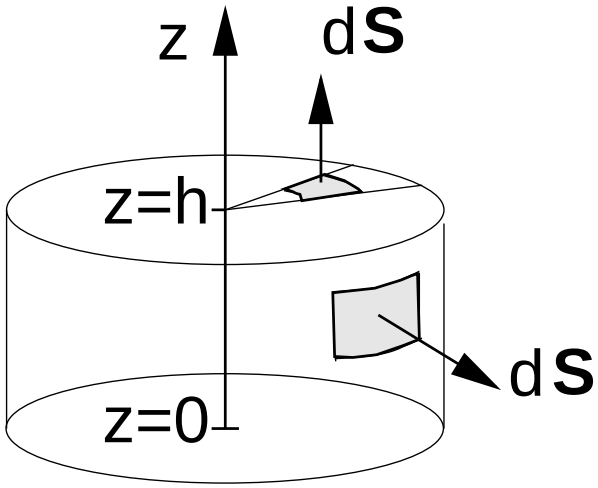
where $\mathbf{c}$ is a constant vector.

♣ Example of extension to Gauss' Theorem

Q $U = x^2 + y^2 + z^2$ is a scalar field, and volume V is the cylinder $x^2 + y^2 \leq a^2$, $0 \leq z \leq h$. Compute the surface integral

$$\int_S U d\mathbf{S}$$

over the surface of the cylinder.



A It is immediately clear from symmetry that there is no contribution from the curved surface of the cylinder since for every vector surface element there exists an equal and opposite element with the same value of U. We therefore need consider only the top and bottom faces.

Top face:

$$U = x^2 + y^2 + z^2 = r^2 + h^2 \text{ and } d\mathbf{S} = r dr d\phi \hat{\boldsymbol{k}}$$

so

$$\int U d\mathbf{S} = \int_{r=0}^{a} (h^2 + r^2) 2\pi r dr \int_{\phi=0}^{2\pi} d\phi \hat{\boldsymbol{k}} = \hat{\boldsymbol{k}}\pi \left[h^2 r^2 + \frac{1}{2} r^4 \right]_0^a = \pi [h^2 a^2 + \frac{1}{2} a^4] \hat{\boldsymbol{k}}$$

Bottom face:

$$U = r^2 \text{ and } d\mathbf{S} = -r\,dr\,d\phi\hat{\boldsymbol{k}}$$

The contribution from this face is thus $-\frac{\pi a^4}{2}\hat{\boldsymbol{k}}$, and the total integral is $\pi h^2 a^2 \hat{\boldsymbol{k}}$.

On the other hand, using Gauss' Theorem we have to compute

$$\int_V \text{grad } U dV$$

In this case, grad $U = 2\mathbf{r}$,

$$2\int_V (x\hat{\boldsymbol{i}} + y\hat{\boldsymbol{k}} + z\hat{\boldsymbol{k}})r\ dr\ dz\ d\phi$$

The integrations over x and y are zero by symmetry, so that the only remaining part is

$$2\int_{z=0}^{h} z dz \int_{r=0}^{a} r\ dr \int_{\phi=0}^{2\pi} d\phi\hat{\boldsymbol{k}} = \pi a^2 h^2 \hat{\boldsymbol{k}}$$

7.5 Stokes' Theorem

Stokes' Theorem relates a line integral around a closed path to a surface integral over what is called a *capping surface* of the path.

Stokes' Theorem states:

$$\oint_C \mathbf{a}\cdot d\mathbf{l} = \int_S \text{curl}\ \ \mathbf{a}\cdot d\mathbf{S}$$

where S is *any* surface capping the curve C.

Why have we used $d\mathbf{l}$ rather than $d\mathbf{r}$, where $\mathbf{r}$ is the position vector?

There is no good reason for this, as $d\mathbf{l} = d\mathbf{r}$. It just seems to be common usage in line integrals!

7.6 Informal proof

You will recall that in Lecture 5 that we defined curl as the circulation per unit area, and showed that

$$\sum_{\text{around elemental loop}} \mathbf{a} \cdot d\mathbf{l} = dC = (\nabla \times \mathbf{a}) \cdot d\mathbf{S} \quad .$$

Now if we add these little loops together, the internal line sections cancel out because the $d\mathbf{l}$'s are in opposite direction but the field $\mathbf{a}$ is not. This gives the larger surface and the larger bounding contour as shown in Fig. 7.4.

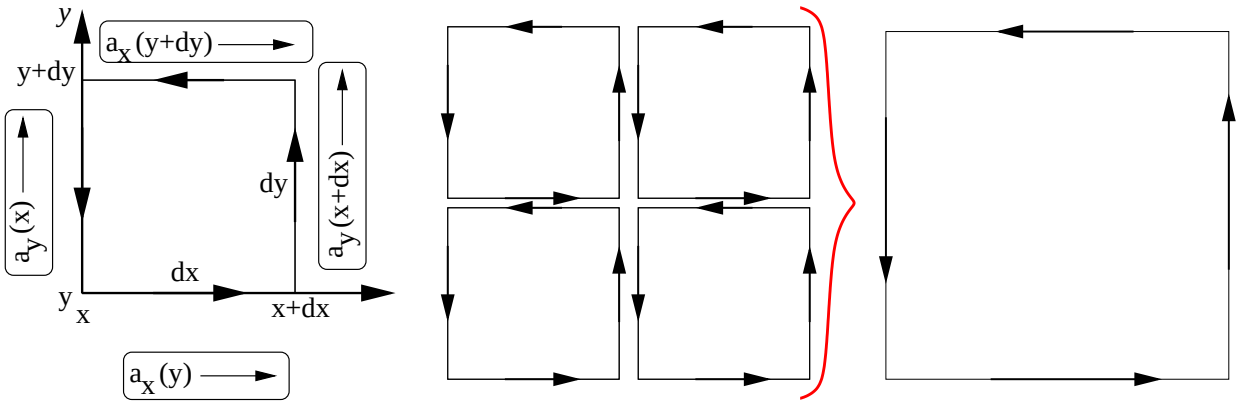


Figure 7.4: An example of an elementary loop, and how they combine together.

For a given contour, the capping surface can be ANY surface bound by the contour. The only requirement is that the surface element vectors point in the "general direction" of a right-handed screw with respect to the sense of the contour integral. See Fig. 7.5.

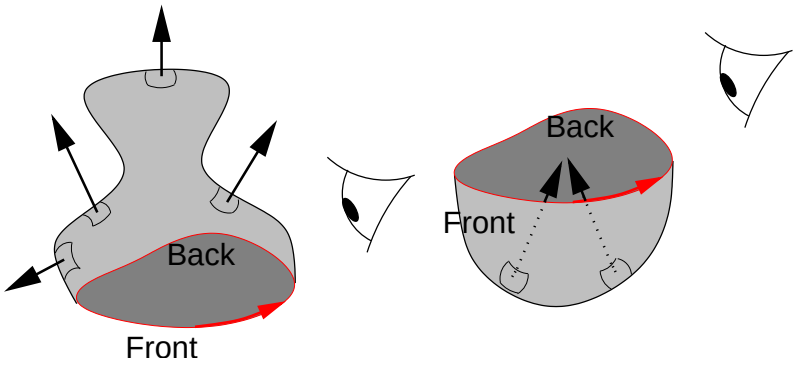


Figure 7.5: For a given contour, the bounding surface can be any shape. $d\mathbf{S}$'s must have a positive component in the sense of a r-h screw wrt the contour sense.

♣ Example of Stokes' Theorem

In practice, (and especially in exam questions!) the bounding contour is often planar, and the capping surface flat or hemispherical or cylindrical.

Q Vector field $\mathbf{a} = x^3\hat{\boldsymbol{j}} - y^3\hat{\boldsymbol{i}}$ and C is the circle of radius R centred on the origin. Derive

$$\oint_C \mathbf{a} \cdot d\mathbf{l}$$

directly and (ii) using Stokes' theorem where the surface is the planar surface bounded by the contour.

A(i) Directly. On the circle of radius R

$$\mathbf{a} = R^3(-\sin^3\theta\hat{\boldsymbol{i}} + \cos^3\theta\hat{\boldsymbol{j}})$$

and

$$d\mathbf{l} = Rd\theta(-\sin\theta\hat{\boldsymbol{i}} + \cos\theta\hat{\boldsymbol{j}})$$

so that:

$$\oint_C \mathbf{a} \cdot d\mathbf{l} = \int_0^{2\pi} R^4(\sin^4\theta + \cos^4\theta)d\theta = \frac{3\pi}{2}R^4,$$

since

$$\int_0^{2\pi} \sin^4\theta d\theta = \int_0^{2\pi} \cos^4\theta d\theta = \frac{3\pi}{4}$$

A(ii) Using Stokes' theorem ...

$$\text{curl } \mathbf{a} = \begin{vmatrix} \hat{\boldsymbol{i}} & \hat{\boldsymbol{j}} & \hat{\boldsymbol{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -y^3 & x^3 & 0 \end{vmatrix} = 3(x^2+y^2)\hat{\boldsymbol{k}} = 3r^2\hat{\boldsymbol{k}}$$

We choose area elements to be circular strips of radius r thickness dr. Then

$$d\mathbf{S} = 2\pi r dr\hat{\boldsymbol{k}} \qquad \text{and} \qquad \int_S \text{curl } \mathbf{a} \cdot d\mathbf{S} = 6\pi \int_0^R r^3 dr = \frac{3\pi}{2}R^4$$

7.7 An Extension to Stokes' Theorem

Just as we considered one extension to Gauss' theorem (not *really* an extension, more of a re-expression), so we will try something similar with Stoke's Theorem.

Again let $\mathbf{a}(\mathbf{r}) = U(\mathbf{r})\mathbf{c}$, where $\mathbf{c}$ is a constant vector. Then

$$\text{curl } \mathbf{a} = U\text{curl } \mathbf{c} + \text{grad } U \times \mathbf{c})$$

Again, curl $\mathbf{c}$ is zero. Stokes' Theorem becomes in this case:

$$\oint_C U(\mathbf{c} \cdot d\mathbf{l}) = \int_S (\text{grad } U \times \mathbf{c} \cdot d\mathbf{S} = \int_S \mathbf{c} \cdot (d\mathbf{S} \times \text{grad } U)$$

or, rearranging the triple scalar products and taking the constant $\mathbf{c}$ out of the integrals gives

$$\mathbf{c} \cdot \oint_C U d\mathbf{l} = -\mathbf{c} \cdot \int_S \text{grad} U \times d\mathbf{S} \quad .$$

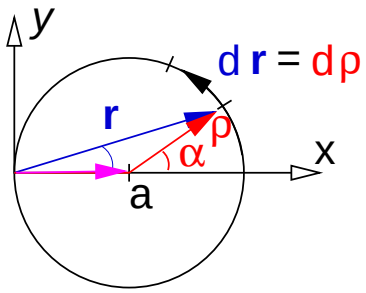
But $\mathbf{c}$ is arbitrary and so

$$\oint_C U d\mathbf{l} = -\int_S \text{grad } U \times d\mathbf{S}$$

7.8 ♣ Example of extension to Stokes' Theorem

Q Derive $\oint_C U d\mathbf{r}$ (i) directly and (ii) using Stokes', where $U = x^2 + y^2 + z^2$ and the line integral is taken around C the circle $(x-a)^2 + y^2 = a^2$ and $z = 0$.

Note that, for no special reason, we have used $d\mathbf{r}$ here not $d\mathbf{l}$.



A(i) First some preamble.

If the circle were centred at the origin, we would write $d\mathbf{r} = ad\theta\hat{\mathbf{e}}_\theta = ad\theta(-\sin\theta\hat{\boldsymbol{\imath}} + \cos\theta\hat{\boldsymbol{\jmath}})$. For such a circle the magnitude $r = |\mathbf{r}| = a$, a constant and so $dr = 0$.

However, in this example $d\mathbf{r}$ is not always in the direction of $\hat{\mathbf{e}}_\theta$, and $dr \neq 0$. Could you write down $d\mathbf{r}$? If not, revise Lecture 3, where we saw that in plane polars $x = r\cos\theta$, $y = r\sin\theta$ and the general expression is

$$d\mathbf{r} = dx\hat{\boldsymbol{\imath}} + dy\hat{\boldsymbol{\jmath}} = (\cos\theta dr - r\sin\theta d\theta)\hat{\boldsymbol{\imath}} + (\sin\theta dr + r\cos\theta d\theta)\hat{\boldsymbol{\jmath}}$$

To avoid having to find an expression for r in terms of θ, we will perform a coordinate transformation by writing $\mathbf{r} = [a, 0]^\top + \boldsymbol{\rho}$. So, $x = (a + \rho\cos\alpha)$ and $y = \rho\sin\alpha$, and on the circle itself where $\rho = a$

$$\mathbf{r} = a(1+\cos\alpha)\hat{\boldsymbol{\imath}} + a\sin\alpha\hat{\boldsymbol{\jmath}} \ ,$$

$$d\mathbf{r} = ad\alpha(-\sin\alpha\hat{\boldsymbol{\imath}} \ + \ \cos\alpha\hat{\boldsymbol{\jmath}}) \ ,$$

and, as $z = 0$ on the circle,

$$U = a^2(1+\cos\alpha)^2 + a^2\sin^2\alpha = 2a^2(1+\cos\alpha) \ .$$

The line integral becomes

$$\oint U d\mathbf{r} = 2a^3 \int_{\alpha=0}^{2\pi} (1+\cos\alpha)(-\sin\alpha\hat{\boldsymbol{\imath}} + \cos\alpha\hat{\boldsymbol{\jmath}}) d\alpha = 2\pi a^3\hat{\boldsymbol{\jmath}}$$

A(ii) Now using Stokes' ...

For a planar surface covering the disc, the surface element can be written using the new parametrization as

$$d\mathbf{S} = \rho \ d\rho \ d\alpha\hat{\boldsymbol{k}}$$

Remember that $U = x^2 + y^2 + z^2 = r^2$, and as $z = 0$ in the plane

$$\text{grad } U = 2(x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}) = 2(a+\rho\cos\alpha)\hat{\boldsymbol{\imath}} + 2\rho\sin\alpha\hat{\boldsymbol{\jmath}} \ .$$

Be careful to note that x, y are specified for any point on the disc, not on its circular boundary!

So

$$d\mathbf{S}\times\text{grad}U = 2\rho \ d\rho \ d\alpha \begin{vmatrix} \hat{\boldsymbol{\imath}} & \hat{\boldsymbol{\jmath}} & \hat{\boldsymbol{k}} \\ 0 & 0 & 1 \\ (a+\rho\cos\alpha) & \rho\sin\alpha & 0 \end{vmatrix} = 2\rho[-\rho\sin\alpha\hat{\boldsymbol{\imath}} + (a+\rho\cos\alpha)\hat{\boldsymbol{\jmath}}] \ d\rho \ d\alpha$$

Both $\int_0^{2\pi}\sin\alpha d\alpha = 0$ and $\int_0^{2\pi}\cos\alpha d\alpha = 0$, so we are left with

$$\int_S d\mathbf{S}\times\text{grad}U = \int_{\rho=0}^{a}\int_{\alpha=0}^{2\pi} 2\rho a\hat{\boldsymbol{\jmath}} \ d\rho \ d\alpha = 2\pi a^3\hat{\boldsymbol{\jmath}}$$

Revised Oct 2013

Lecture 8

Engineering applications

In Lecture 6 we saw one classic example of the application of vector calculus to Maxwell's equation.

In this lecture we explore a few more examples from fluid mechanics and heat transfer. As with Maxwell's eqations, the examples show how vector calculus provides a powerful way of representing underlying physics.

The power come from the fact that div, grad and curl have a significance or meaning which is more immediate than a collection of partial derivatives. Vector calculus will, with practice, become a convenient shorthand for you.

- Electricity – Ampère's Law
- Fluid Mechanics - The Continuity Equation
- Thermo: The Heat Conduction Equation
- Mechanics/Electrostatics - Conservative fields
- The Inverse Square Law of force
- Gravitational field due to distributed mass
- Gravitational field inside body
- Pressure forces in non-uniform flows

8.1 Electricity – Ampère's Law

If the frequency is low, the displacement current in Maxwell's equation curl$\mathbf{H} = \mathbf{J} + \partial\mathbf{D}/\partial t$ is negligible, and we find

$$\text{curl}\mathbf{H} = \mathbf{J}$$

Hence

$$\int_S \text{curl}\mathbf{H} \cdot d\mathbf{S} = \int_S \mathbf{J} \cdot d\mathbf{S}$$

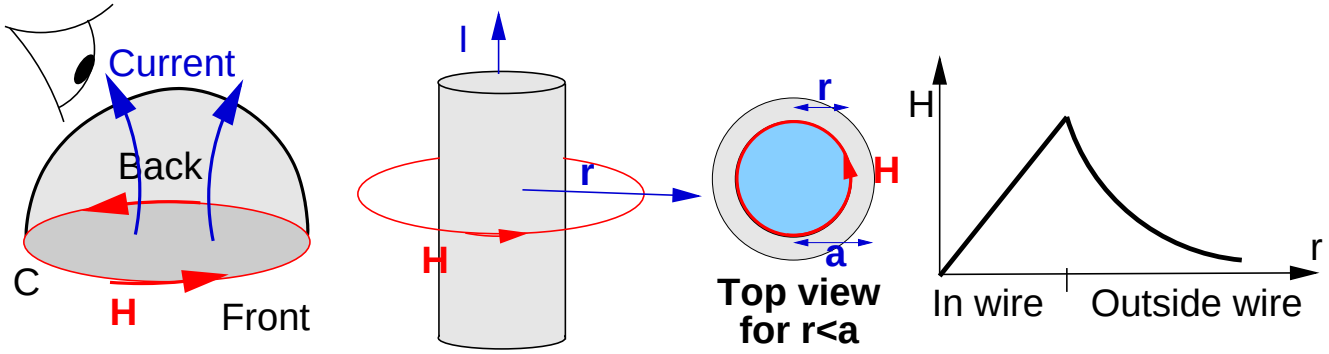
or

$$\oint \mathbf{H} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot d\mathbf{S}$$

where $\int_S \mathbf{J} \cdot d\mathbf{S}$ is total current through the surface.

Now consider the $\mathbf{H}$ around a straight wire carrying current I. Symmetry tells us the $\mathbf{H}$ is in the $\hat{\mathbf{e}}_\theta$ direction, in a rhs screw sense with respect to the current. (You might check this against Biot-Savart's law.)

Suppose we asked what is the magnitude of $\mathbf{H}$?



Inside the wire, the bounding contour only encloses a fraction $(\pi r^2)/(\pi a^2)$ of the current, and so

$$\begin{aligned} H2\pi r &= \int \mathbf{J} \cdot d\mathbf{S} = I(r^2/A^2) \\ \Rightarrow\ H &= Ir/2\pi A^2 \end{aligned}$$

whereas outside we enclose all the current, and so

$$\begin{aligned} H2\pi r &= \int \mathbf{J} \cdot d\mathbf{S} = I \\ \Rightarrow\ H &= I/2\pi r \end{aligned}$$

A plot is shown in the Figure.

8.2 Fluid Mechanics - The Continuity Equation

The **Continuity Equation** expresses the condition of conservation of mass in a fluid flow. The continuity principle applied to *any* volume (called a *control volume*) may be expressed in words as follows:

> "The net rate of mass flow of fluid out of the control volume must equal the rate of decrease of the mass of fluid within the control volume"

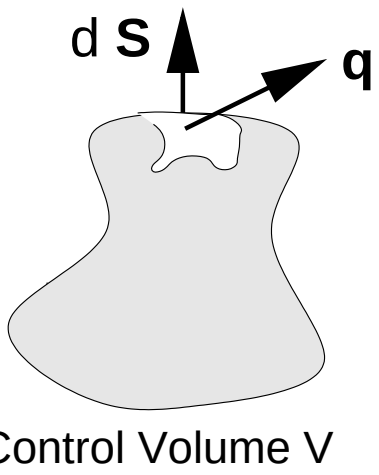


Figure 8.1:

To express the above as a mathematical equation, we denote the velocity of the fluid at each point of the flow by $\mathbf{q}(\mathbf{r})$ (a vector field) and the density by $\rho(\mathbf{r})$ (a scalar field). The element of rate-of-volume-loss through surface $d\mathbf{S}$ is $d\dot{V} = \mathbf{q}\cdot d\mathbf{S}$, so the rate of mass loss is

$$d\dot{M} = \rho\mathbf{q}\cdot d\mathbf{S},$$

so that the total rate of mass loss from the volume is

$$-\frac{\partial}{\partial t}\int_V \rho(\mathbf{r})dV = \int_S \rho\mathbf{q}\cdot d\mathbf{S}.$$

Assuming that the volume of interest is fixed, this is the same as

$$-\int_V \frac{\partial\rho}{\partial t}dV = \int_S \rho\mathbf{q}\cdot d\mathbf{S}\ .$$

Now we use Gauss' Theorem to transform the RHS into a volume integral

$$-\int_V \frac{\partial\rho}{\partial t}dV = \int_V \text{div}\ (\rho\mathbf{q})dV\ .$$

The two volume integrals can be equal for any control volume V only if the two integrands are equal at each point of the flow. This leads to the mathematical formulation of

The Continuity Equation:

$$\text{div } (\rho \mathbf{q}) = -\frac{\partial \rho}{\partial t}$$

Notice that if the density doesn't vary with time, div $(\rho \mathbf{q}) = 0$, and if the density doesn't vary with position then

The Continuity Equation for uniform, time-invariant density:

$$\text{div } (\mathbf{q}) = 0 \quad .$$

In this last case, we can say that the flow $\mathbf{q}$ is solenoidal.

8.3 Thermodynamics - The Heat Conduction Equation

Flow of heat is very similar to flow of fluid, and heat flow satisfies a similar continuity equation. The flow is characterized by the heat current density $\mathbf{q}(\mathbf{r})$ (heat flow per unit area and time), sometimes misleadingly called heat flux.

Assuming that there is no mass flow across the boundary of the control volume and no source of heat inside it, the rate of flow of heat out of the control volume by conduction must equal the rate of decrease of internal energy (constant volume) or enthalpy (constant pressure) within it. This leads to the equation

$$\text{div } \mathbf{q} = -\rho c \frac{\partial T}{\partial t},$$

where ρ is the density of the conducting medium, c its specific heat (both are assumed constant) and T is the temperature.

In order to solve for the temperature field another equation is required, linking q to the temperature gradient. This is

$$\mathbf{q} = -\kappa \text{grad } T,$$

where κ is the thermal conductivity of the medium. Combining the two equations gives the *heat conduction equation*:

$$-\text{div } \mathbf{q} = \kappa \text{div grad } T = \kappa \nabla^2 T = \rho c \frac{\partial T}{\partial t}$$

where it has been assumed that κ is a constant. In steady flow the temperature field satisfies Laplace's Equation $\nabla^2 T = 0$.

8.4 Mechanics - Conservative fields of force

A conservative field of force is one for which the work done

$$\int_A^B \mathbf{F} \cdot d\mathbf{r},$$

moving from A to B is indep. of path taken. As we saw in Lecture 4, conservative fields must satisfy the condition

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = 0,$$

Stokes' tells us that this is

$$\int_S \text{curl } \mathbf{F} \cdot d\mathbf{S} = 0,$$

where S is *any* surface bounded by C.

But if true for *any* C containing A and B, it must be that

$$\text{curl } \mathbf{F} = \mathbf{0}$$

Conservative fields are irrotational
All radial fields are irrotational

One way (actually the only way) of satisfying this condition is for

$$\mathbf{F} = \boldsymbol{\nabla}\, U$$

The scalar field $U(\mathbf{r})$ is the Potential Function

8.5 The Inverse Square Law of force

Radial forces are found in electrostatics and gravitation — so they are certainly irrotational and conservative.

But in nature these radial forces are also inverse square laws. One reason why this may be so is that it turns out to be the only central force field which is **solenoidal**, i.e. has zero divergence.

If $\mathbf{F} = f(r)\mathbf{r}$,

$$\text{div } \mathbf{F} = 3f(r) + rf'(r).$$

For div $\mathbf{F} = 0$ we conclude

$$r\frac{df}{dr} + 3f = 0$$

or

$$\frac{df}{f} + 3\frac{dr}{r} = 0.$$

Integrating with respect to r gives $fr^3 = const = A$, so that

$$\mathbf{F} = \frac{A\mathbf{r}}{r^3}, \quad |\mathbf{F}| = \frac{A}{r^2}.$$

The condition of zero divergence of the inverse square force field applies everywhere except at $\mathbf{r} = \mathbf{0}$, where the divergence is infinite.

To show this, calculate the outward normal flux out of a sphere of radius R centered on the origin when $\mathbf{F} = F\hat{\mathbf{r}}$. This is

$$\int_S \mathbf{F} \cdot d\mathbf{S} = F \int_{\text{Sphere}} \hat{\mathbf{r}} \cdot d\mathbf{S} = F \int_{\text{Sphere}} d = F4\pi R^2 = 4\pi A = \text{Constant}.$$

Gauss tells us that this flux must be equal to

$$\int_V \text{div } \mathbf{F} dV = \int_0^R \text{div } \mathbf{F} 4\pi r^2 dr$$

where we have done the volume integral as a summation over thin shells of surface area $4\pi r^2$ and thickness dr.

But for all finite r, div$\mathbf{F} = 0$, so div$\mathbf{F}$ must be infinite at the origin.

The flux integral is thus

- zero — for any volume which does not contain the origin
- $4\pi A$ for any volume which does contain it.

8.6 Gravitational field due to distributed mass: Poisson's Equation

If one tried the same approach as §8.4 for the gravitational field, $A = Gm$, where m is the mass at the origin and G the universal gravitational constant, one would run into the problem that there is no such thing as point mass.

We can make progress though by considering distributed mass.

The mass contained in each small volume element dV is ρdV and this will make a contribution $-4\pi\rho G dV$ to the flux integral from the control volume. Mass outside the control volume makes no contribution, so that we obtain the equation

$$\int_S \mathbf{F} \cdot d\mathbf{S} = -4\pi G \int_V \rho dV.$$

Transforming the left hand integral by Gauss' Theorem gives

$$\int_V \text{div } \mathbf{F} dV = -4\pi G \int_V \rho dV$$

which, since it is true for any V, implies that

$$-\text{div } \mathbf{F} = 4\pi\rho G.$$

Since the gravitational field is also conservative (i.e. irrotational) it must have an associated potential function U, so that $\mathbf{F} = \text{grad } U$. It follows that the gravitational potential U satisfies

Poisson's Equation

$$\nabla^2 U = 4\pi\rho G \ .$$

Using the integral form of Poisson's equation, it is possible to calculate the gravitational field inside a spherical body whose density is a function of radius only. We have

$$4\pi R^2 F = 4\pi G \int_0^R 4\pi r^2 \rho dr,$$

where $F = |\mathbf{F}|$, or

$$|F| = \frac{G}{R^2} \int_0^R 4\pi r^2 \rho dr = \frac{MG}{R^2},$$

where M is the total mass inside radius R. For the case of uniform density, this is equal to $M = \frac{4}{3}\pi\rho R^3$ and $|F| = \frac{4}{3}\pi\rho G R$.

8.7 Pressure forces in non-uniform flows

When a body is immersed in a flow it experiences a net pressure force

$$\mathbf{F}_p = -\int_S p d\mathbf{S},$$

where S is the surface of the body. If the pressure p is non-uniform, this integral is not zero. The integral can be transformed using Gauss' Theorem to give the alternative expression

$$\mathbf{F}_p = -\int_V \text{grad } p \; dV,$$

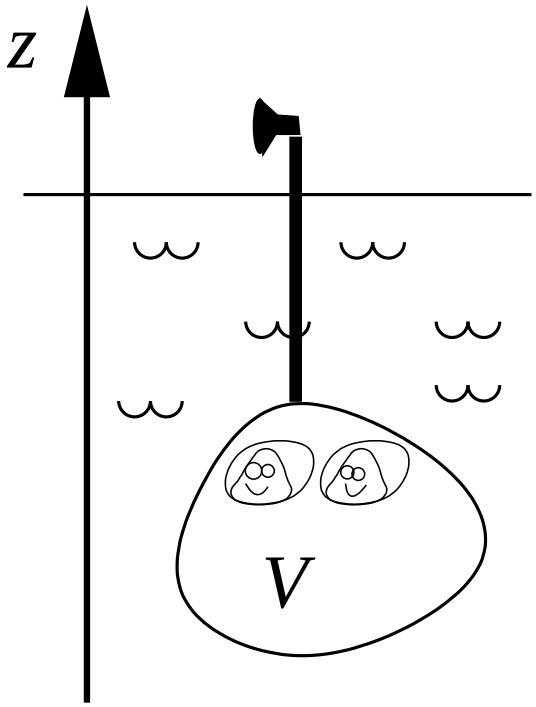
where V is the volume of the body. In the simple hydrostatic case $p + \rho g z = constant$, so that

$$\text{grad } p = -\rho g k$$

and the net pressure force is simply

$$\mathbf{F}_p = g\hat{\mathbf{k}} \int_V \rho dV$$

which, in agreement with Archimedes' principle, is equal to the weight of fluid displaced.



Chapter Six: Vectors in Euclidean Space

1 Vectors in Euclidean Space

1.1 Introduction

In single-variable calculus, the functions that one encounters are functions of a variable (usually x or t) that varies over some subset of the real number line (which we denote by $\mathbb{R}$). For such a function, say, $y = f(x)$, the **graph** of the function f consists of the points $(x, y) = (x, f(x))$. These points lie in the **Euclidean plane**, which, in the **Cartesian** or **rectangular** coordinate system, consists of all ordered pairs of real numbers (a, b). We use the word "Euclidean" to denote a system in which all the usual rules of Euclidean geometry hold. We denote the Euclidean plane by $\mathbb{R}^2$; the "2" represents the number of *dimensions* of the plane. The Euclidean plane has two perpendicular **coordinate axes**: the x-axis and the y-axis.

In vector (or multivariable) calculus, we will deal with functions of two or three variables (usually x, y or x, y, z, respectively). The graph of a function of two variables, say, $z = f(x, y)$, lies in **Euclidean space**, which in the Cartesian coordinate system consists of all ordered triples of real numbers (a, b, c). Since Euclidean space is 3-dimensional, we denote it by $\mathbb{R}^3$. The graph of f consists of the points $(x, y, z) = (x, y, f(x, y))$. The 3-dimensional coordinate system of Euclidean space can be represented on a flat surface, such as this page or a blackboard, only by giving the illusion of three dimensions, in the manner shown in Figure 1.1.1. Euclidean space has three mutually perpendicular coordinate axes (x, y and z), and three mutually perpendicular coordinate planes: the xy-plane, yz-plane and xz-plane (see Figure 1.1.2).

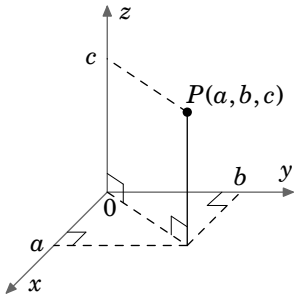


Figure 1.1.1

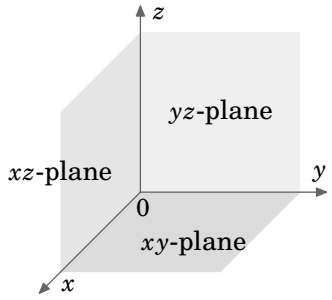


Figure 1.1.2

The coordinate system shown in Figure 1.1.1 is known as a **right-handed coordinate system**, because it is possible, using the right hand, to point the index finger in the positive direction of the x-axis, the middle finger in the positive direction of the y-axis, and the thumb in the positive direction of the z-axis, as in Figure 1.1.3.

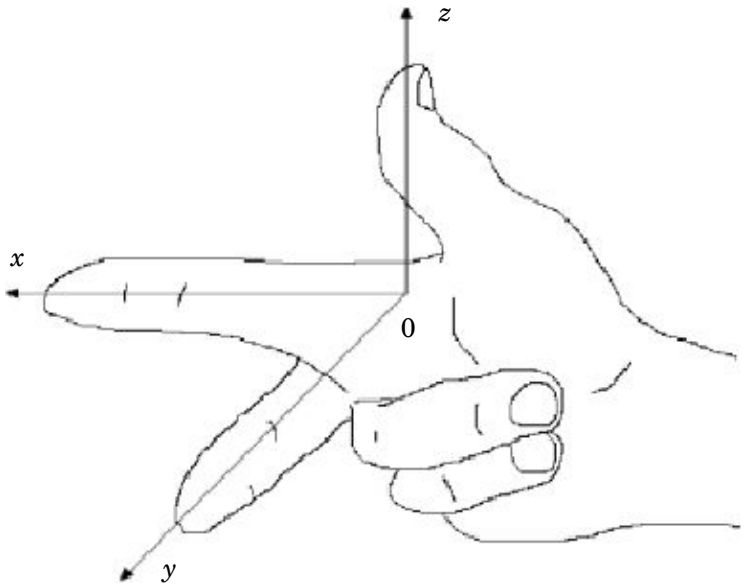


Figure 1.1.3 Right-handed coordinate system

An equivalent way of defining a right-handed system is if you can point your thumb upwards in the positive z-axis direction while using the remaining four fingers to rotate the x-axis towards the y-axis. Doing the same thing with the left hand is what defines a **left-handed coordinate system**. Notice that switching the x- and y-axes in a right-handed system results in a left-handed system, and that rotating either type of system does not change its "handedness". Throughout the book we will use a right-handed system.

For functions of three variables, the graphs exist in 4-dimensional space (i.e. $\mathbb{R}^4$), which we can not see in our 3-dimensional space, let alone simulate in 2-dimensional space. So we can only think of 4-dimensional space abstractly. For an entertaining discussion of this subject, see the book by ABBOTT.[1]

So far, we have discussed the *position* of an object in 2-dimensional or 3-dimensional space. But what about something such as the velocity of the object, or its acceleration? Or the gravitational force acting on the object? These phenomena all seem to involve motion and *direction* in some way. This is where the idea of a *vector* comes in.

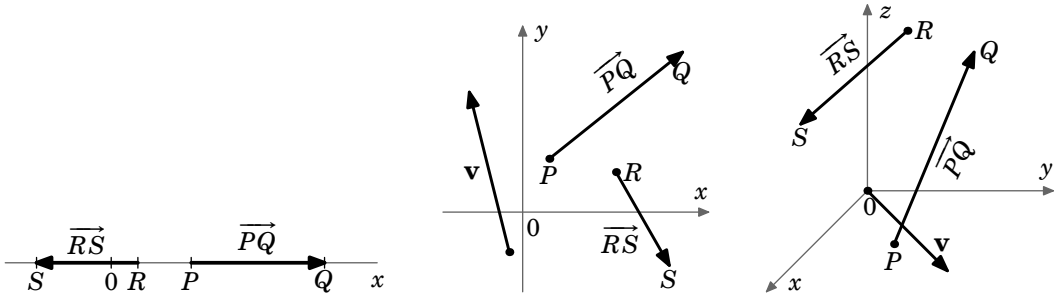
[1] One thing you will learn is why a 4-dimensional creature would be able to reach inside an egg and remove the yolk without cracking the shell!

You have already dealt with velocity and acceleration in single-variable calculus. For example, for motion along a straight line, if $y = f(t)$ gives the displacement of an object after time t, then $dy/dt = f'(t)$ is the velocity of the object at time t. The derivative $f'(t)$ is just a number, which is positive if the object is moving in an agreed-upon "positive" direction, and negative if it moves in the opposite of that direction. So you can think of that number, which was called the velocity of the object, as having two components: a *magnitude*, indicated by a nonnegative number, preceded by a *direction*, indicated by a plus or minus symbol (representing motion in the positive direction or the negative direction, respectively), i.e. $f'(t) = \pm a$ for some number $a \geq 0$. Then a is the magnitude of the velocity (normally called the *speed* of the object), and the $\pm$ represents the direction of the velocity (though the $+$ is usually omitted for the positive direction).

For motion along a straight line, i.e. in a 1-dimensional space, the velocities are also contained in that 1-dimensional space, since they are just numbers. For general motion along a curve in 2- or 3-dimensional space, however, velocity will need to be represented by a multi-dimensional object which should have both a magnitude and a direction. A geometric object which has those features is an arrow, which in elementary geometry is called a "directed line segment". This is the motivation for how we will define a vector.

Definition 1.1. A (nonzero) **vector** is a directed line segment drawn from a point P (called its **initial point**) to a point Q (called its **terminal point**), with P and Q being distinct points. The vector is denoted by $\overrightarrow{PQ}$. Its **magnitude** is the length of the line segment, denoted by $\|\overrightarrow{PQ}\|$, and its **direction** is the same as that of the directed line segment. The **zero vector** is just a point, and it is denoted by $\mathbf{0}$.

To indicate the direction of a vector, we draw an arrow from its initial point to its terminal point. We will often denote a vector by a single bold-faced letter (e.g. $\mathbf{v}$) and use the terms "magnitude" and "length" interchangeably. Note that our definition could apply to systems with any number of dimensions (see Figure 1.1.4 (a)-(c)).



(a) One dimension **(b)** Two dimensions **(c)** Three dimensions

Figure 1.1.4 Vectors in different dimensions

A few things need to be noted about the zero vector. Our motivation for what a vector is included the notions of magnitude and direction. What is the magnitude of the zero vector? We define it to be zero, i.e. $\|\mathbf{0}\| = 0$. This agrees with the definition of the zero vector as just a point, which has zero length. What about the direction of the zero vector? A single point really has no well-defined direction. Notice that we were careful to only define the direction of a *nonzero* vector, which is well-defined since the initial and terminal points are distinct. Not everyone agrees on the direction of the zero vector. Some contend that the zero vector has *arbitrary* direction (i.e. can take any direction), some say that it has *indeterminate* direction (i.e. the direction can not be determined), while others say that it has *no* direction. Our definition of the zero vector, however, does not require it to have a direction, and we will leave it at that.[2]

Now that we know what a vector is, we need a way of determining when two vectors are equal. This leads us to the following definition.

Definition 1.2. Two nonzero vectors are **equal** if they have the same magnitude and the same direction. Any vector with zero magnitude is equal to the zero vector.

By this definition, vectors with the same magnitude and direction but with different initial points would be equal. For example, in Figure 1.1.5 the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ all have the same magnitude $\sqrt{5}$ (by the Pythagorean Theorem). And we see that $\mathbf{u}$ and $\mathbf{w}$ are parallel, since they lie on lines having the same slope $\frac{1}{2}$, and they point in the same direction. So $\mathbf{u} = \mathbf{w}$, even though they have different initial points. We also see that $\mathbf{v}$ is parallel to $\mathbf{u}$ but points in the opposite direction. So $\mathbf{u} \neq \mathbf{v}$.

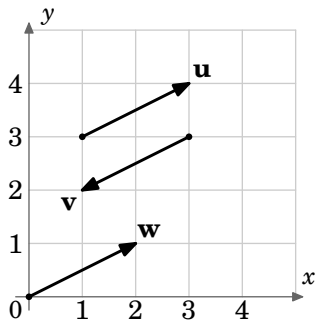


Figure 1.1.5

So we can see that there are an infinite number of vectors for a given magnitude and direction, those vectors all being equal and differing only by their initial and terminal points. Is there a single vector which we can choose to represent all those equal vectors? The answer is yes, and is suggested by the vector $\mathbf{w}$ in Figure 1.1.5.

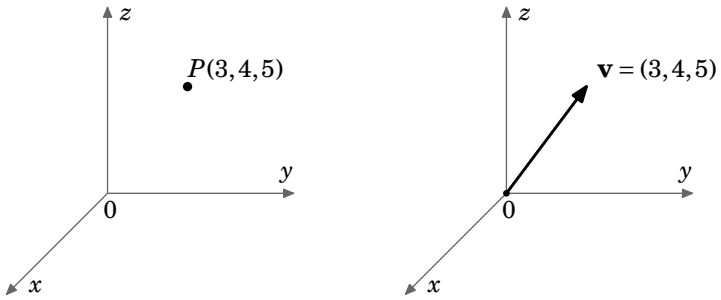
[2] In the subject of linear algebra there is a more abstract way of defining a vector where the concept of "direction" is not really used. See ANTON and RORRES.

Unless otherwise indicated, when speaking of "the vector" with a given magnitude and direction, we will mean the one whose initial point is at the origin of the coordinate system.

Thinking of vectors as starting from the origin provides a way of dealing with vectors in a standard way, since every coordinate system has an origin. But there will be times when it is convenient to consider a different initial point for a vector (for example, when adding vectors, which we will do in the next section).

Another advantage of using the origin as the initial point is that it provides an easy correspondence between a vector and its terminal point.

Example 1.1. Let $\mathbf{v}$ be the vector in $\mathbb{R}^3$ whose initial point is at the origin and whose terminal point is $(3,4,5)$. Though the *point* $(3,4,5)$ and the vector $\mathbf{v}$ are different objects, it is convenient to write $\mathbf{v} = (3,4,5)$. When doing this, it is understood that the initial point of $\mathbf{v}$ is at the origin $(0,0,0)$ and the terminal point is $(3,4,5)$.



(a) The point (3,4,5) **(b)** The vector (3,4,5)

Figure 1.1.6 Correspondence between points and vectors

Unless otherwise stated, when we refer to vectors as $\mathbf{v} = (a,b)$ in $\mathbb{R}^2$ or $\mathbf{v} = (a,b,c)$ in $\mathbb{R}^3$, we mean vectors in Cartesian coordinates starting at the origin. Also, we will write the zero vector $\mathbf{0}$ in $\mathbb{R}^2$ and $\mathbb{R}^3$ as $(0,0)$ and $(0,0,0)$, respectively.

The point-vector correspondence provides an easy way to check if two vectors are equal, without having to determine their magnitude and direction. Similar to seeing if two points are the same, you are now seeing if the terminal points of vectors starting at the origin are the same. For each vector, find the (unique!) vector it equals whose initial point is the origin. Then compare the coordinates of the terminal points of these "new" vectors: if those coordinates are the same, then the original vectors are equal. To get the "new" vectors starting at the origin, you *translate* each vector to start at the origin by subtracting the coordinates of the original initial point from the original terminal point. The resulting point will be the terminal point of the "new" vector whose initial point is the origin. Do this for each original vector then compare.

Example 1.2. Consider the vectors $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ in $\mathbb{R}^3$, where $P=(2,1,5), Q=(3,5,7), R=(1,-3,-2)$ and $S=(2,1,0)$. Does $\overrightarrow{PQ}=\overrightarrow{RS}$?

Solution: The vector $\overrightarrow{PQ}$ is equal to the vector $\mathbf{v}$ with initial point $(0,0,0)$ and terminal point $Q-P=(3,5,7)-(2,1,5)=(3-2,5-1,7-5)=(1,4,2)$.

Similarly, $\overrightarrow{RS}$ is equal to the vector $\mathbf{w}$ with initial point $(0,0,0)$ and terminal point $S-R=(2,1,0)-(1,-3,-2)=(2-1,1-(-3),0-(-2))=(1,4,2)$.

So $\overrightarrow{PQ}=\mathbf{v}=(1,4,2)$ and $\overrightarrow{RS}=\mathbf{w}=(1,4,2)$.

$\therefore \overrightarrow{PQ}=\overrightarrow{RS}$

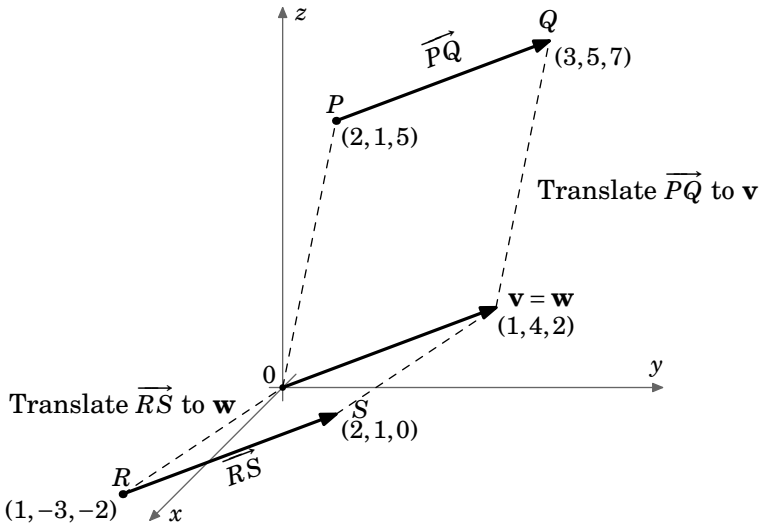


Figure 1.1.7

Recall the distance formula for points in the Euclidean plane:

For points $P=(x_1,y_1)$, $Q=(x_2,y_2)$ in $\mathbb{R}^2$, the distance d between P and Q is:

$$d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \tag{1.1}$$

By this formula, we have the following result:

For a vector $\overrightarrow{PQ}$ in $\mathbb{R}^2$ with initial point $P=(x_1,y_1)$ and terminal point $Q=(x_2,y_2)$, the magnitude of $\overrightarrow{PQ}$ is:

$$\left\|\overrightarrow{PQ}\right\|=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \tag{1.2}$$

Finding the magnitude of a vector $\mathbf{v} = (a,b)$ in $\mathbb{R}^2$ is a special case of formula (1.2) with $P = (0,0)$ and $Q = (a,b)$:

For a vector $\mathbf{v} = (a,b)$ in $\mathbb{R}^2$, the magnitude of $\mathbf{v}$ is:

$$\|\mathbf{v}\| = \sqrt{a^2 + b^2} \tag{1.3}$$

To calculate the magnitude of vectors in $\mathbb{R}^3$, we need a distance formula for points in Euclidean space (we will postpone the proof until the next section):

Theorem 1.1. The distance d between points $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$ in $\mathbb{R}^3$ is:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{1.4}$$

The proof will use the following result:

Theorem 1.2. For a vector $\mathbf{v} = (a,b,c)$ in $\mathbb{R}^3$, the magnitude of $\mathbf{v}$ is:

$$\|\mathbf{v}\| = \sqrt{a^2 + b^2 + c^2} \tag{1.5}$$

Proof: There are four cases to consider:
Case 1: $a = b = c = 0$. Then $\mathbf{v} = \mathbf{0}$, so $\|\mathbf{v}\| = 0 = \sqrt{0^2 + 0^2 + 0^2} = \sqrt{a^2 + b^2 + c^2}$.

Case 2: exactly two of a, b, c *are* 0. Without loss of generality, we assume that $a = b = 0$ and $c \neq 0$ (the other two possibilities are handled in a similar manner). Then $\mathbf{v} = (0,0,c)$, which is a vector of length $|c|$ along the z-axis. So $\|\mathbf{v}\| = |c| = \sqrt{c^2} = \sqrt{0^2 + 0^2 + c^2} = \sqrt{a^2 + b^2 + c^2}$.

Case 3: exactly one of a, b, c *is* 0. Without loss of generality, we assume that $a = 0$, $b \neq 0$ and $c \neq 0$ (the other two possibilities are handled in a similar manner). Then $\mathbf{v} = (0,b,c)$, which is a vector in the yz-plane, so by the Pythagorean Theorem we have $\|\mathbf{v}\| = \sqrt{b^2 + c^2} = \sqrt{0^2 + b^2 + c^2} = \sqrt{a^2 + b^2 + c^2}$.

Case 4: none of a, b, c *are* 0. Without loss of generality, we can assume that a, b, c are all positive (the other seven possibilities are handled in a similar manner). Consider the points $P = (0,0,0)$, $Q = (a,b,c)$, $R = (a,b,0)$, and $S = (a,0,0)$, as shown in Figure 1.1.8. Applying the Pythagorean Theorem to the right triangle $\triangle PSR$ gives $|PR|^2 = a^2 + b^2$. A second application of the Pythagorean Theorem, this time to the right triangle $\triangle PQR$, gives $\|\mathbf{v}\| = |PQ| = \sqrt{|PR|^2 + |QR|^2} = \sqrt{a^2 + b^2 + c^2}$.
This proves the theorem. **QED**

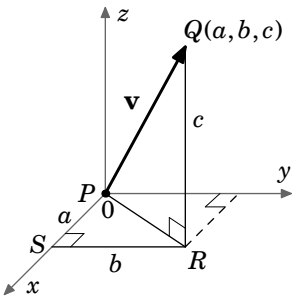


Figure 1.1.8

Example 1.3. Calculate the following:

(a) The magnitude of the vector $\overrightarrow{PQ}$ in $\mathbb{R}^2$ with $P=(-1,2)$ and $Q=(5,5)$.
Solution: By formula (1.2), $\left\|\overrightarrow{PQ}\right\| = \sqrt{(5-(-1))^2+(5-2)^2} = \sqrt{36+9} = \sqrt{45} = 3\sqrt{5}$.

(b) The magnitude of the vector $\mathbf{v}=(8,3)$ in $\mathbb{R}^2$.
Solution: By formula (1.3), $\|\mathbf{v}\| = \sqrt{8^2+3^2} = \sqrt{73}$.

(c) The distance between the points $P=(2,-1,4)$ and $Q=(4,2,-3)$ in $\mathbb{R}^2$.
Solution: By formula (1.4), the distance $d = \sqrt{(4-2)^2+(2-(-1))^2+(-3-4)^2} = \sqrt{4+9+49} = \sqrt{62}$.

(d) The magnitude of the vector $\mathbf{v}=(5,8,-2)$ in $\mathbb{R}^3$.
Solution: By formula (1.5), $\|\mathbf{v}\| = \sqrt{5^2+8^2+(-2)^2} = \sqrt{25+64+4} = \sqrt{93}$.

Exercises

A

1. Calculate the magnitudes of the following vectors:
(a) $\mathbf{v}=(2,-1)$ (b) $\mathbf{v}=(2,-1,0)$ (c) $\mathbf{v}=(3,2,-2)$ (d) $\mathbf{v}=(0,0,1)$ (e) $\mathbf{v}=(6,4,-4)$

2. For the points $P=(1,-1,1)$, $Q=(2,-2,2)$, $R=(2,0,1)$, $S=(3,-1,2)$, does $\overrightarrow{PQ}=\overrightarrow{RS}$?

3. For the points $P=(0,0,0)$, $Q=(1,3,2)$, $R=(1,0,1)$, $S=(2,3,4)$, does $\overrightarrow{PQ}=\overrightarrow{RS}$?

B

4. Let $\mathbf{v}=(1,0,0)$ and $\mathbf{w}=(a,0,0)$ be vectors in $\mathbb{R}^3$. Show that $\|\mathbf{w}\| = |a|\,\|\mathbf{v}\|$.

5. Let $\mathbf{v}=(a,b,c)$ and $\mathbf{w}=(3a,3b,3c)$ be vectors in $\mathbb{R}^3$. Show that $\|\mathbf{w}\| = 3\,\|\mathbf{v}\|$.

C

6. Though we will see a simple proof of Theorem 1.1 in the next section, it is possible to prove it using methods similar to those in the proof of Theorem 1.2. Prove the special case of Theorem 1.1 where the points $P=(x_1,y_1,z_1)$ and $Q=(x_2,y_2,z_2)$ satisfy the following conditions:
$x_2 > x_1 > 0$, $y_2 > y_1 > 0$, and $z_2 > z_1 > 0$.
(*Hint: Think of Case 4 in the proof of Theorem 1.2, and consider Figure 1.1.9.*)

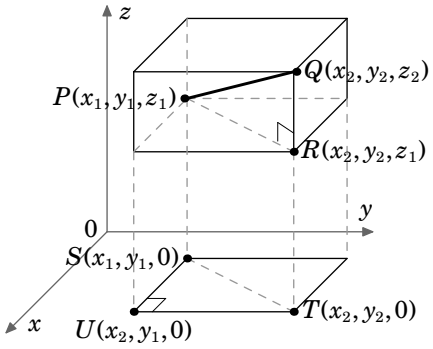


Figure 1.1.9

1.2 Vector Algebra

Now that we know what vectors are, we can start to perform some of the usual algebraic operations on them (e.g. addition, subtraction). Before doing that, we will introduce the notion of a *scalar*.

Definition 1.3. A **scalar** is a quantity that can be represented by a single number.

For our purposes, scalars will always be real numbers.[3] Examples of scalar quantities are mass, electric charge, and speed (not velocity).[4] We can now define *scalar multiplication* of a vector.

Definition 1.4. For a scalar k and a nonzero vector $\mathbf{v}$, the **scalar multiple** of $\mathbf{v}$ by k, denoted by $k\mathbf{v}$, is the vector whose magnitude is $|k|\|\mathbf{v}\|$, points in the same direction as $\mathbf{v}$ if $k > 0$, points in the opposite direction as $\mathbf{v}$ if $k < 0$, and is the zero vector $\mathbf{0}$ if $k = 0$. For the zero vector $\mathbf{0}$, we define $k\mathbf{0} = \mathbf{0}$ for any scalar k.

Two vectors $\mathbf{v}$ and $\mathbf{w}$ are **parallel** (denoted by $\mathbf{v} \parallel \mathbf{w}$) if one is a scalar multiple of the other. You can think of scalar multiplication of a vector as stretching or shrinking the vector, and as flipping the vector in the opposite direction if the scalar is a negative number (see Figure 1.2.1).

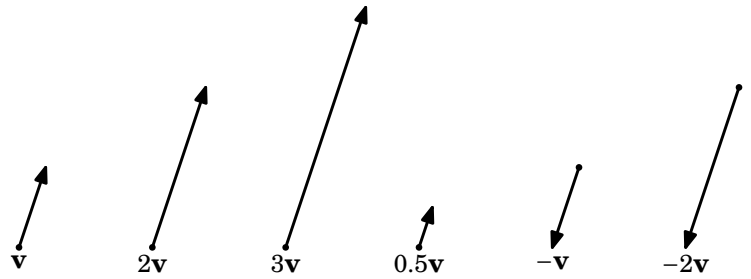


Figure 1.2.1

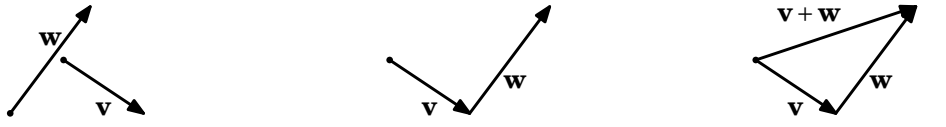
Recall that **translating** a nonzero vector means that the initial point of the vector is changed but the magnitude and direction are preserved. We are now ready to define the *sum* of two vectors.

Definition 1.5. The **sum** of vectors $\mathbf{v}$ and $\mathbf{w}$, denoted by $\mathbf{v} + \mathbf{w}$, is obtained by translating $\mathbf{w}$ so that its initial point is at the terminal point of $\mathbf{v}$; the initial point of $\mathbf{v} + \mathbf{w}$ is the initial point of $\mathbf{v}$, and its terminal point is the new terminal point of $\mathbf{w}$.

[3] The term *scalar* was invented by 19$^{\text{th}}$ century Irish mathematician, physicist and astronomer William Rowan Hamilton, to convey the sense of something that could be represented by a point on a scale or graduated ruler. The word vector comes from Latin, where it means "carrier".

[4] An alternate definition of scalars and vectors, used in physics, is that under certain types of coordinate transformations (e.g. rotations), a quantity that is not affected is a scalar, while a quantity that is affected (in a certain way) is a vector. See MARION for details.

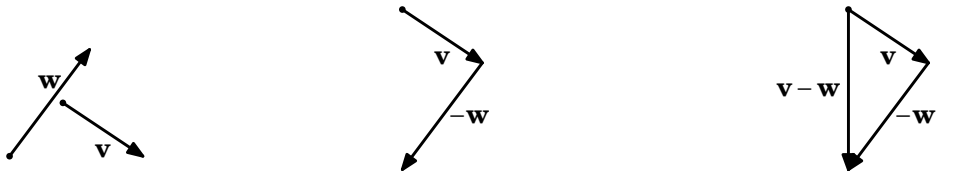
Intuitively, adding $\mathbf{w}$ to $\mathbf{v}$ means tacking on $\mathbf{w}$ to the end of $\mathbf{v}$ (see Figure 1.2.2).



(a) Vectors $\mathbf{v}$ and $\mathbf{w}$ **(b)** Translate $\mathbf{w}$ to the end of $\mathbf{v}$ **(c)** The sum $\mathbf{v}+\mathbf{w}$

Figure 1.2.2 Adding vectors $\mathbf{v}$ and $\mathbf{w}$

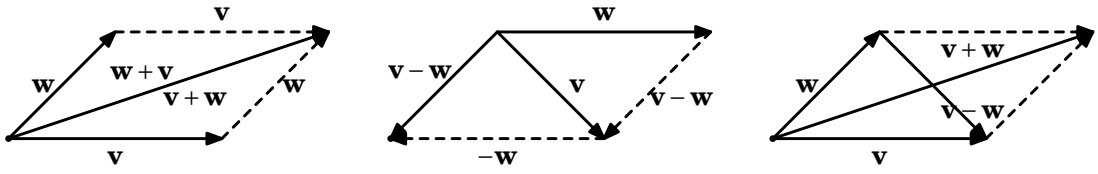
Notice that our definition is valid for the zero vector (which is just a point, and hence can be translated), and so we see that $\mathbf{v}+\mathbf{0}=\mathbf{v}=\mathbf{0}+\mathbf{v}$ for any vector $\mathbf{v}$. In particular, $\mathbf{0}+\mathbf{0}=\mathbf{0}$. Also, it is easy to see that $\mathbf{v}+(-\mathbf{v})=\mathbf{0}$, as we would expect. In general, since the scalar multiple $-\mathbf{v}=-1\mathbf{v}$ is a well-defined vector, we can define **vector subtraction** as follows: $\mathbf{v}-\mathbf{w}=\mathbf{v}+(-\mathbf{w})$. See Figure 1.2.3.



(a) Vectors $\mathbf{v}$ and $\mathbf{w}$ **(b)** Translate $-\mathbf{w}$ to the end of $\mathbf{v}$ **(c)** The difference $\mathbf{v}-\mathbf{w}$

Figure 1.2.3 Subtracting vectors $\mathbf{v}$ and $\mathbf{w}$

Figure 1.2.4 shows the use of "geometric proofs" of various laws of vector algebra, that is, it uses laws from elementary geometry to prove statements about vectors. For example, (a) shows that $\mathbf{v}+\mathbf{w}=\mathbf{w}+\mathbf{v}$ for any vectors $\mathbf{v}$, $\mathbf{w}$. And (c) shows how you can think of $\mathbf{v}-\mathbf{w}$ as the vector that is tacked on to the end of $\mathbf{w}$ to add up to $\mathbf{v}$.



(a) Add vectors **(b)** Subtract vectors **(c)** Combined add/subtract

Figure 1.2.4 "Geometric" vector algebra

Notice that we have temporarily abandoned the practice of starting vectors at the origin. In fact, we have not even mentioned coordinates in this section so far. Since we will deal mostly with Cartesian coordinates in this book, the following two theorems are useful for performing vector algebra on vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ starting at the origin.

Theorem 1.3. Let $\mathbf{v} = (v_1, v_2)$, $\mathbf{w} = (w_1, w_2)$ be vectors in $\mathbb{R}^2$, and let k be a scalar. Then
(a) $k\mathbf{v} = (kv_1, kv_2)$
(b) $\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2)$

Proof: (a) Without loss of generality, we assume that $v_1, v_2 > 0$ (the other possibilities are handled in a similar manner). If $k = 0$ then $k\mathbf{v} = 0\mathbf{v} = \mathbf{0} = (0,0) = (0v_1, 0v_2) = (kv_1, kv_2)$, which is what we needed to show. If $k \neq 0$, then (kv_1, kv_2) lies on a line with slope $\frac{kv_2}{kv_1} = \frac{v_2}{v_1}$, which is the same as the slope of the line on which $\mathbf{v}$ (and hence $k\mathbf{v}$) lies, and (kv_1, kv_2) points in the same direction on that line as $k\mathbf{v}$. Also, by formula (1.3) the magnitude of (kv_1, kv_2) is $\sqrt{(kv_1)^2 + (kv_2)^2} = \sqrt{k^2v_1^2 + k^2v_2^2} = \sqrt{k^2(v_1^2 + v_2^2)} = |k|\sqrt{v_1^2 + v_2^2} = |k|\,\|\mathbf{v}\|$. So $k\mathbf{v}$ and (kv_1, kv_2) have the same magnitude and direction. This proves (a).

(b) Without loss of generality, we assume that $v_1, v_2, w_1, w_2 > 0$ (the other possibilities are handled in a similar manner). From Figure 1.2.5, we see that when translating $\mathbf{w}$ to start at the end of $\mathbf{v}$, the new terminal point of $\mathbf{w}$ is $(v_1 + w_1, v_2 + w_2)$, so by the definition of $\mathbf{v} + \mathbf{w}$ this must be the terminal point of $\mathbf{v} + \mathbf{w}$. This proves (b). **QED**

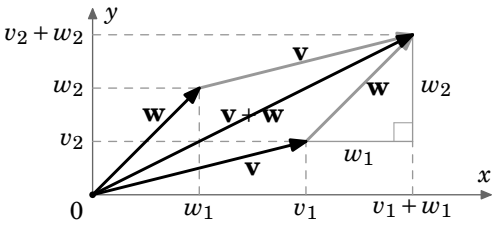


Figure 1.2.5

Theorem 1.4. Let $\mathbf{v} = (v_1, v_2, v_3)$, $\mathbf{w} = (w_1, w_2, w_3)$ be vectors in $\mathbb{R}^3$, let k be a scalar. Then
(a) $k\mathbf{v} = (kv_1, kv_2, kv_3)$
(b) $\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, v_3 + w_3)$

The following theorem summarizes the basic laws of vector algebra.

Theorem 1.5. For any vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and scalars k, l, we have

(a) $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$	Commutative Law
(b) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$	Associative Law
(c) $\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v}$	Additive Identity
(d) $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$	Additive Inverse
(e) $k(l\mathbf{v}) = (kl)\mathbf{v}$	Associative Law
(f) $k(\mathbf{v} + \mathbf{w}) = k\mathbf{v} + k\mathbf{w}$	Distributive Law
(g) $(k + l)\mathbf{v} = k\mathbf{v} + l\mathbf{v}$	Distributive Law

Proof: (a) We already presented a geometric proof of this in Figure 1.2.4(a).

(b) To illustrate the difference between analytic proofs and geometric proofs in vector algebra, we will present both types here. For the analytic proof, we will use vectors in $\mathbb{R}^3$ (the proof for $\mathbb{R}^2$ is similar).

Let $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, $\mathbf{w} = (w_1, w_2, w_3)$ be vectors in $\mathbb{R}^3$. Then

$$\begin{aligned}
\mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (u_1, u_2, u_3) + ((v_1, v_2, v_3) + (w_1, w_2, w_3)) \\
&= (u_1, u_2, u_3) + (v_1 + w_1, v_2 + w_2, v_3 + w_3) && \text{by Theorem 1.4(b)} \\
&= (u_1 + (v_1 + w_1), u_2 + (v_2 + w_2), u_3 + (v_3 + w_3)) && \text{by Theorem 1.4(b)} \\
&= ((u_1 + v_1) + w_1, (u_2 + v_2) + w_2, (u_3 + v_3) + w_3) && \text{by properties of real numbers} \\
&= (u_1 + v_1, u_2 + v_2, u_3 + v_3) + (w_1, w_2, w_3) && \text{by Theorem 1.4(b)} \\
&= (\mathbf{u} + \mathbf{v}) + \mathbf{w}
\end{aligned}$$

This completes the analytic proof of (b). Figure 1.2.6 provides the geometric proof.

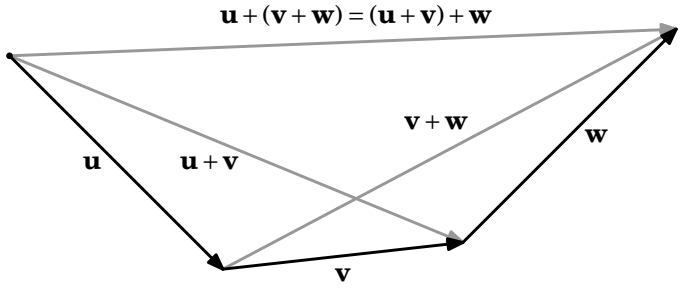


Figure 1.2.6 Associative Law for vector addition

(c) We already discussed this on p.10.

(d) We already discussed this on p.10.

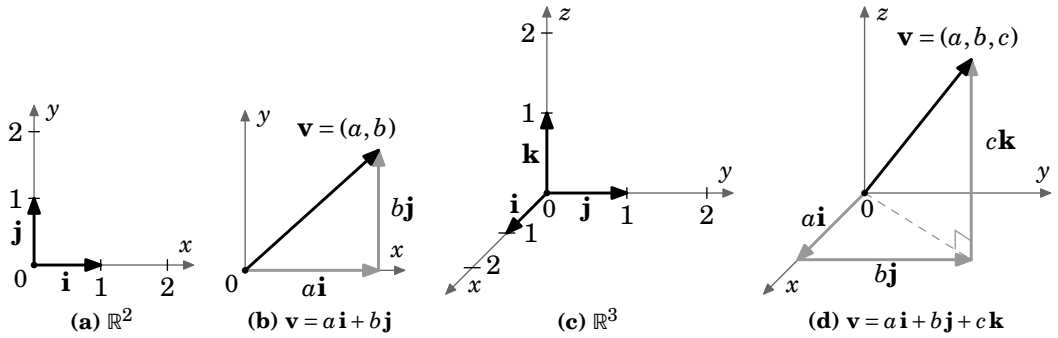
(e) We will prove this for a vector $\mathbf{v} = (v_1, v_2, v_3)$ in $\mathbb{R}^3$ (the proof for $\mathbb{R}^2$ is similar):

$$\begin{aligned}
k(l\mathbf{v}) &= k(lv_1, lv_2, lv_3) && \text{by Theorem 1.4(a)} \\
&= (klv_1, klv_2, klv_3) && \text{by Theorem 1.4(a)} \\
&= (kl)(v_1, v_2, v_3) && \text{by Theorem 1.4(a)} \\
&= (kl)\mathbf{v}
\end{aligned}$$

(f) and (g): Left as exercises for the reader. **QED**

A **unit vector** is a vector with magnitude 1. Notice that for any nonzero vector $\mathbf{v}$, the vector $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ is a unit vector which points in the same direction as $\mathbf{v}$, since $\frac{1}{\|\mathbf{v}\|} > 0$ and $\left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| = \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|} = 1$. Dividing a nonzero vector $\mathbf{v}$ by $\|\mathbf{v}\|$ is often called *normalizing* $\mathbf{v}$.

There are specific unit vectors which we will often use, called the **basis vectors**: $\mathbf{i} = (1,0,0)$, $\mathbf{j} = (0,1,0)$, and $\mathbf{k} = (0,0,1)$ in $\mathbb{R}^3$; $\mathbf{i} = (1,0)$ and $\mathbf{j} = (0,1)$ in $\mathbb{R}^2$.
These are useful for several reasons: they are mutually perpendicular, since they lie on distinct coordinate axes; they are all unit vectors: $\|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1$; every vector can be written as a unique scalar combination of the basis vectors: $\mathbf{v} = (a,b) = a\,\mathbf{i} + b\,\mathbf{j}$ in $\mathbb{R}^2$, $\mathbf{v} = (a,b,c) = a\,\mathbf{i} + b\,\mathbf{j} + c\,\mathbf{k}$ in $\mathbb{R}^3$. See Figure 1.2.7.



(a) $\mathbb{R}^2$ **(b)** $\mathbf{v} = a\,\mathbf{i} + b\,\mathbf{j}$ **(c)** $\mathbb{R}^3$ **(d)** $\mathbf{v} = a\,\mathbf{i} + b\,\mathbf{j} + c\,\mathbf{k}$

Figure 1.2.7 Basis vectors in different dimensions

When a vector $\mathbf{v} = (a, b, c)$ is written as $\mathbf{v} = a\,\mathbf{i} + b\,\mathbf{j} + c\,\mathbf{k}$, we say that $\mathbf{v}$ is in *component form*, and that a, b, and c are the ***i, j, and k components***, respectively, of $\mathbf{v}$. We have:

$$\mathbf{v} = v_1\,\mathbf{i} + v_2\,\mathbf{j} + v_3\,\mathbf{k}, k \text{ a scalar} \Longrightarrow k\mathbf{v} = kv_1\,\mathbf{i} + kv_2\,\mathbf{j} + kv_3\,\mathbf{k}$$

$$\mathbf{v} = v_1\,\mathbf{i} + v_2\,\mathbf{j} + v_3\,\mathbf{k}, \mathbf{w} = w_1\,\mathbf{i} + w_2\,\mathbf{j} + w_3\,\mathbf{k} \Longrightarrow \mathbf{v} + \mathbf{w} = (v_1 + w_1)\mathbf{i} + (v_2 + w_2)\mathbf{j} + (v_3 + w_3)\mathbf{k}$$

$$\mathbf{v} = v_1\,\mathbf{i} + v_2\,\mathbf{j} + v_3\,\mathbf{k} \Longrightarrow \|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$$

Example 1.4. Let $\mathbf{v} = (2, 1, -1)$ and $\mathbf{w} = (3, -4, 2)$ in $\mathbb{R}^3$.

(a) Find $\mathbf{v} - \mathbf{w}$.
Solution: $\mathbf{v} - \mathbf{w} = (2 - 3, 1 - (-4), -1 - 2) = (-1, 5, -3)$

(b) Find $3\mathbf{v} + 2\mathbf{w}$.
Solution: $3\mathbf{v} + 2\mathbf{w} = (6, 3, -3) + (6, -8, 4) = (12, -5, 1)$

(c) Write $\mathbf{v}$ and $\mathbf{w}$ in component form.
Solution: $\mathbf{v} = 2\,\mathbf{i} + \mathbf{j} - \mathbf{k}$, $\mathbf{w} = 3\,\mathbf{i} - 4\,\mathbf{j} + 2\,\mathbf{k}$

(d) Find the vector $\mathbf{u}$ such that $\mathbf{u} + \mathbf{v} = \mathbf{w}$.
Solution: By Theorem 1.5, $\mathbf{u} = \mathbf{w} - \mathbf{v} = -(\mathbf{v} - \mathbf{w}) = -(-1, 5, -3) = (1, -5, 3)$, by part(a).

(e) Find the vector $\mathbf{u}$ such that $\mathbf{u} + \mathbf{v} + \mathbf{w} = \mathbf{0}$.
Solution: By Theorem 1.5, $\mathbf{u} = -\mathbf{w} - \mathbf{v} = -(3, -4, 2) - (2, 1, -1) = (-5, 3, -1)$.

(f) Find the vector $\mathbf{u}$ such that $2\mathbf{u} + \mathbf{i} - 2\mathbf{j} = \mathbf{k}$.
Solution: $2\mathbf{u} = -\mathbf{i} + 2\mathbf{j} + \mathbf{k} \Longrightarrow \mathbf{u} = -\frac{1}{2}\,\mathbf{i} + \mathbf{j} + \frac{1}{2}\,\mathbf{k}$

(g) Find the unit vector $\frac{\mathbf{v}}{\|\mathbf{v}\|}$.
Solution: $\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{\sqrt{2^2 + 1^2 + (-1)^2}}(2, 1, -1) = \left(\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}\right)$

We can now easily prove Theorem 1.1 from the previous section. The distance d between two points $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$ in $\mathbb{R}^3$ is the same as the length of the vector $\mathbf{w} - \mathbf{v}$, where the vectors $\mathbf{v}$ and $\mathbf{w}$ are defined as $\mathbf{v} = (x_1, y_1, z_1)$ and $\mathbf{w} = (x_2, y_2, z_2)$ (see Figure 1.2.8). So since $\mathbf{w} - \mathbf{v} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$, then $d = \|\mathbf{w} - \mathbf{v}\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$ by Theorem 1.2.

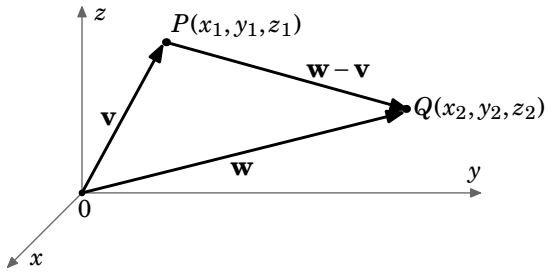


Figure 1.2.8 Proof of Theorem 1.2: $d = \|\mathbf{w} - \mathbf{v}\|$

Exercises

A

1. Let $\mathbf{v} = (-1, 5, -2)$ and $\mathbf{w} = (3, 1, 1)$.

(a) Find $\mathbf{v} - \mathbf{w}$. (b) Find $\mathbf{v} + \mathbf{w}$. (c) Find $\frac{\mathbf{v}}{\|\mathbf{v}\|}$. (d) Find $\left\| \frac{1}{2}(\mathbf{v} - \mathbf{w}) \right\|$.

(e) Find $\left\| \frac{1}{2}(\mathbf{v} + \mathbf{w}) \right\|$. (f) Find $-2\mathbf{v} + 4\mathbf{w}$. (g) Find $\mathbf{v} - 2\mathbf{w}$.

(h) Find the vector $\mathbf{u}$ such that $\mathbf{u} + \mathbf{v} + \mathbf{w} = \mathbf{i}$.

(i) Find the vector $\mathbf{u}$ such that $\mathbf{u} + \mathbf{v} + \mathbf{w} = 2\mathbf{j} + \mathbf{k}$.

(j) Is there a scalar m such that $m(\mathbf{v} + 2\mathbf{w}) = \mathbf{k}$? If so, find it.

2. For the vectors $\mathbf{v}$ and $\mathbf{w}$ from Exercise 1, is $\|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v}\| - \|\mathbf{w}\|$? If not, which quantity is larger?

3. For the vectors $\mathbf{v}$ and $\mathbf{w}$ from Exercise 1, is $\|\mathbf{v} + \mathbf{w}\| = \|\mathbf{v}\| + \|\mathbf{w}\|$? If not, which quantity is larger?

B

4. Prove Theorem 1.5(f) for $\mathbb{R}^3$.

5. Prove Theorem 1.5(g) for $\mathbb{R}^3$.

C

6. We know that every vector in $\mathbb{R}^3$ can be written as a scalar combination of the vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$. Can every vector in $\mathbb{R}^3$ be written as a scalar combination of just $\mathbf{i}$ and $\mathbf{j}$, i.e. for any vector $\mathbf{v}$ in $\mathbb{R}^3$, are there scalars m, n such that $\mathbf{v} = m\mathbf{i} + n\mathbf{j}$? Justify your answer.

1.3 Dot Product

You may have noticed that while we did define multiplication of a vector by a scalar in the previous section on vector algebra, we did not define multiplication of a vector by a vector. We will now see one type of multiplication of vectors, called the *dot product*.

Definition 1.6. Let $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ be vectors in $\mathbb{R}^3$.
The **dot product** of $\mathbf{v}$ and $\mathbf{w}$, denoted by $\mathbf{v} \cdot \mathbf{w}$, is given by:

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + v_3 w_3 \tag{1.6}$$

Similarly, for vectors $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$ in $\mathbb{R}^2$, the dot product is:

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 \tag{1.7}$$

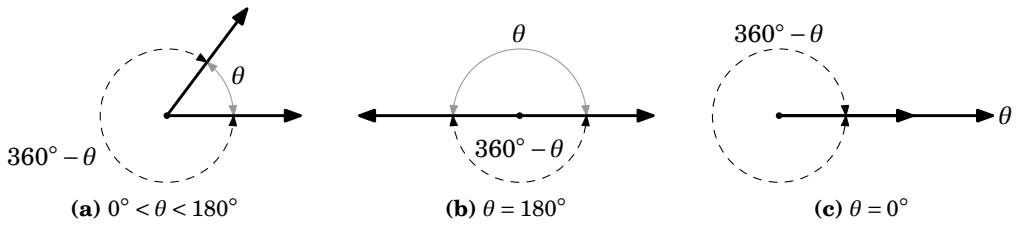
Notice that the dot product of two vectors is a scalar, not a vector. So the associative law that holds for multiplication of numbers and for addition of vectors (see Theorem 1.5(b),(e)), does *not* hold for the dot product of vectors. Why? Because for vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, the dot product $\mathbf{u} \cdot \mathbf{v}$ is a scalar, and so $(\mathbf{u} \cdot \mathbf{v}) \cdot \mathbf{w}$ is not defined since the left side of that dot product (the part in parentheses) is a scalar and not a vector.

For vectors $\mathbf{v} = v_1 \mathbf{i} + v_2 \mathbf{j} + v_3 \mathbf{k}$ and $\mathbf{w} = w_1 \mathbf{i} + w_2 \mathbf{j} + w_3 \mathbf{k}$ in component form, the dot product is still $\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + v_3 w_3$.

Also notice that we defined the dot product in an analytic way, i.e. by referencing vector coordinates. There is a geometric way of defining the dot product, which we will now develop as a consequence of the analytic definition.

Definition 1.7. The **angle** between two nonzero vectors with the same initial point is the smallest angle between them.

We do not define the angle between the zero vector and any other vector. Any two nonzero vectors with the same initial point have two angles between them: θ and $360° - \theta$. We will always choose the smallest nonnegative angle θ between them, so that $0° \le \theta \le 180°$. See Figure 1.3.1.



(a) $0° < \theta < 180°$ **(b)** $\theta = 180°$ **(c)** $\theta = 0°$

Figure 1.3.1 Angle between vectors

We can now take a more geometric view of the dot product by establishing a relationship between the dot product of two vectors and the angle between them.

Theorem 1.6. Let $\mathbf{v}$, $\mathbf{w}$ be nonzero vectors, and let θ be the angle between them. Then

$$\cos\theta = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\,\|\mathbf{w}\|} \tag{1.8}$$

Proof: We will prove the theorem for vectors in $\mathbb{R}^3$ (the proof for $\mathbb{R}^2$ is similar). Let $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$. By the Law of Cosines (see Figure 1.3.2), we have

$$\|\mathbf{v}-\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\|\,\|\mathbf{w}\|\cos\theta \tag{1.9}$$

(note that equation (1.9) holds even for the "degenerate" cases $\theta = 0°$ and $180°$).

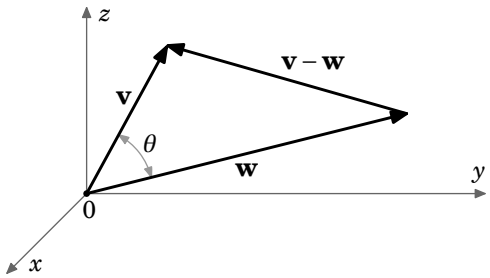


Figure 1.3.2

Since $\mathbf{v}-\mathbf{w} = (v_1-w_1, v_2-w_2, v_3-w_3)$, expanding $\|\mathbf{v}-\mathbf{w}\|^2$ in equation (1.9) gives

$$\begin{aligned}
\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\|\,\|\mathbf{w}\|\cos\theta &= (v_1-w_1)^2 + (v_2-w_2)^2 + (v_3-w_3)^2 \\
&= (v_1^2 - 2v_1w_1 + w_1^2) + (v_2^2 - 2v_2w_2 + w_2^2) + (v_3^2 - 2v_3w_3 + w_3^2) \\
&= (v_1^2 + v_2^2 + v_3^2) + (w_1^2 + w_2^2 + w_3^2) - 2(v_1w_1 + v_2w_2 + v_3w_3) \\
&= \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2(\mathbf{v}\cdot\mathbf{w})\text{ , so} \\
-2\|\mathbf{v}\|\,\|\mathbf{w}\|\cos\theta &= -2(\mathbf{v}\cdot\mathbf{w})\text{ , so since } \mathbf{v}\neq\mathbf{0} \text{ and } \mathbf{w}\neq\mathbf{0} \text{ then} \\
\cos\theta &= \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\,\|\mathbf{w}\|}\text{ , since } \|\mathbf{v}\|>0 \text{ and } \|\mathbf{w}\|>0. \quad \textbf{QED}
\end{aligned}$$

Example 1.5. Find the angle θ between the vectors $\mathbf{v} = (2,1,-1)$ and $\mathbf{w} = (3,-4,1)$.

Solution: Since $\mathbf{v}\cdot\mathbf{w} = (2)(3)+(1)(-4)+(-1)(1) = 1$, $\|\mathbf{v}\| = \sqrt{6}$, and $\|\mathbf{w}\| = \sqrt{26}$, then

$$\cos\theta = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\,\|\mathbf{w}\|} = \frac{1}{\sqrt{6}\sqrt{26}} = \frac{1}{2\sqrt{39}} \approx 0.08 \implies \theta = 85.41°$$

Two nonzero vectors are **perpendicular** if the angle between them is $90°$. Since $\cos 90° = 0$, we have the following important corollary to Theorem 1.6:

Corollary 1.7. Two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ are perpendicular if and only if $\mathbf{v}\cdot\mathbf{w} = 0$.

We will write $\mathbf{v}\perp\mathbf{w}$ to indicate that $\mathbf{v}$ and $\mathbf{w}$ are perpendicular.

Since $\cos\theta > 0$ for $0° \le \theta < 90°$ and $\cos\theta < 0$ for $90° < \theta \le 180°$, we also have:

Corollary 1.8. If θ is the angle between nonzero vectors $\mathbf{v}$ and $\mathbf{w}$, then

$$\mathbf{v}\cdot\mathbf{w} \text{ is } \begin{cases} > 0 & \text{for } 0° \le \theta < 90° \\ 0 & \text{for } \theta = 90° \\ < 0 & \text{for } 90° < \theta \le 180° \end{cases}$$

By Corollary 1.8, the dot product can be thought of as a way of telling if the angle between two vectors is acute, obtuse, or a right angle, depending on whether the dot product is positive, negative, or zero, respectively. See Figure 1.3.3.

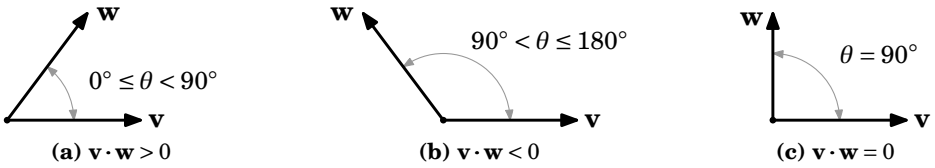


Figure 1.3.3 Sign of the dot product & angle between vectors

Example 1.6. Are the vectors $\mathbf{v} = (-1, 5, -2)$ and $\mathbf{w} = (3, 1, 1)$ perpendicular?

Solution: Yes, $\mathbf{v} \perp \mathbf{w}$ since $\mathbf{v}\cdot\mathbf{w} = (-1)(3) + (5)(1) + (-2)(1) = 0$.

The following theorem summarizes the basic properties of the dot product.

Theorem 1.9. For any vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and scalar k, we have

(a) $\mathbf{v}\cdot\mathbf{w} = \mathbf{w}\cdot\mathbf{v}$	Commutative Law
(b) $(k\mathbf{v})\cdot\mathbf{w} = \mathbf{v}\cdot(k\mathbf{w}) = k(\mathbf{v}\cdot\mathbf{w})$	Associative Law
(c) $\mathbf{v}\cdot\mathbf{0} = 0 = \mathbf{0}\cdot\mathbf{v}$	
(d) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v} + \mathbf{u}\cdot\mathbf{w}$	Distributive Law
(e) $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w}$	Distributive Law
(f) $\lvert\mathbf{v}\cdot\mathbf{w}\rvert \le \lVert\mathbf{v}\rVert\,\lVert\mathbf{w}\rVert$	Cauchy-Schwarz Inequality[5]

Proof: The proofs of parts (a)-(e) are straightforward applications of the definition of the dot product, and are left to the reader as exercises. We will prove part (f).
(f) If either $\mathbf{v} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$, then $\mathbf{v}\cdot\mathbf{w} = 0$ by part (c), and so the inequality holds trivially. So assume that $\mathbf{v}$ and $\mathbf{w}$ are nonzero vectors. Then by Theorem 1.6,

$$\begin{aligned} \mathbf{v}\cdot\mathbf{w} &= \cos\theta\,\lVert\mathbf{v}\rVert\,\lVert\mathbf{w}\rVert \text{ , so} \\ \lvert\mathbf{v}\cdot\mathbf{w}\rvert &= \lvert\cos\theta\rvert\,\lVert\mathbf{v}\rVert\,\lVert\mathbf{w}\rVert \text{ , so} \\ \lvert\mathbf{v}\cdot\mathbf{w}\rvert &\le \lVert\mathbf{v}\rVert\,\lVert\mathbf{w}\rVert \text{ since } \lvert\cos\theta\rvert \le 1. \end{aligned}$$

QED

[5] Also known as the Cauchy-Schwarz-Buniakovski Inequality.

Using Theorem 1.9, we see that if $\mathbf{u}\cdot\mathbf{v}=0$ and $\mathbf{u}\cdot\mathbf{w}=0$, then $\mathbf{u}\cdot(k\mathbf{v}+l\mathbf{w})=k(\mathbf{u}\cdot\mathbf{v})+l(\mathbf{u}\cdot\mathbf{w})=k(0)+l(0)=0$ for all scalars k,l. Thus, we have the following fact:

If $\mathbf{u}\perp\mathbf{v}$ and $\mathbf{u}\perp\mathbf{w}$, then $\mathbf{u}\perp(k\mathbf{v}+l\mathbf{w})$ for all scalars k,l.

For vectors $\mathbf{v}$ and $\mathbf{w}$, the collection of all scalar combinations $k\mathbf{v}+l\mathbf{w}$ is called the **span** of $\mathbf{v}$ and $\mathbf{w}$. If nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ are parallel, then their span is a line; if they are not parallel, then their span is a plane. So what we showed above is that a vector which is perpendicular to two other vectors is also perpendicular to their span.

The dot product can be used to derive properties of the magnitudes of vectors, the most important of which is the *Triangle Inequality*, as given in the following theorem:

Theorem 1.10. For any vectors $\mathbf{v}$, $\mathbf{w}$, we have
(a) $\|\mathbf{v}\|^2=\mathbf{v}\cdot\mathbf{v}$
(b) $\|\mathbf{v}+\mathbf{w}\|\le\|\mathbf{v}\|+\|\mathbf{w}\|$ Triangle Inequality
(c) $\|\mathbf{v}-\mathbf{w}\|\ge\|\mathbf{v}\|-\|\mathbf{w}\|$

Proof: (a) Left as an exercise for the reader.

(b) By part (a) and Theorem 1.9, we have

$$\begin{aligned}\|\mathbf{v}+\mathbf{w}\|^2&=(\mathbf{v}+\mathbf{w})\cdot(\mathbf{v}+\mathbf{w})=\mathbf{v}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{w}+\mathbf{w}\cdot\mathbf{v}+\mathbf{w}\cdot\mathbf{w}\\&=\|\mathbf{v}\|^2+2(\mathbf{v}\cdot\mathbf{w})+\|\mathbf{w}\|^2\text{ , so since }a\le|a|\text{ for any real number }a\text{, we have}\\&\le\|\mathbf{v}\|^2+2|\mathbf{v}\cdot\mathbf{w}|+\|\mathbf{w}\|^2\text{ , so by Theorem 1.9(f) we have}\\&\le\|\mathbf{v}\|^2+2\|\mathbf{v}\|\,\|\mathbf{w}\|+\|\mathbf{w}\|^2=(\|\mathbf{v}\|+\|\mathbf{w}\|)^2\text{ and so}\\\|\mathbf{v}+\mathbf{w}\|&\le\|\mathbf{v}\|+\|\mathbf{w}\|\text{ after taking square roots of both sides, which proves (b).}\end{aligned}$$

(c) Since $\mathbf{v}=\mathbf{w}+(\mathbf{v}-\mathbf{w})$, then $\|\mathbf{v}\|=\|\mathbf{w}+(\mathbf{v}-\mathbf{w})\|\le\|\mathbf{w}\|+\|\mathbf{v}-\mathbf{w}\|$ by the Triangle Inequality, so subtracting $\|\mathbf{w}\|$ from both sides gives $\|\mathbf{v}\|-\|\mathbf{w}\|\le\|\mathbf{v}-\mathbf{w}\|$. **QED**

The Triangle Inequality gets its name from the fact that in any triangle, no one side is longer than the sum of the lengths of the other two sides (see Figure 1.3.4). Another way of saying this is with the familiar statement "the shortest distance between two points is a straight line."

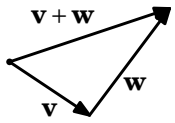


Figure 1.3.4

Exercises

A

1. Let $\mathbf{v}=(5,1,-2)$ and $\mathbf{w}=(4,-4,3)$. Calculate $\mathbf{v}\cdot\mathbf{w}$.

2. Let $\mathbf{v}=-3\mathbf{i}-2\mathbf{j}-\mathbf{k}$ and $\mathbf{w}=6\mathbf{i}+4\mathbf{j}+2\mathbf{k}$. Calculate $\mathbf{v}\cdot\mathbf{w}$.

For Exercises 3-8, find the angle θ between the vectors $\mathbf{v}$ and $\mathbf{w}$.

3. $\mathbf{v} = (5,1,-2)$, $\mathbf{w} = (4,-4,3)$

4. $\mathbf{v} = (7,2,-10)$, $\mathbf{w} = (2,6,4)$

5. $\mathbf{v} = (2,1,4)$, $\mathbf{w} = (1,-2,0)$

6. $\mathbf{v} = (4,2,-1)$, $\mathbf{w} = (8,4,-2)$

7. $\mathbf{v} = -\mathbf{i} + 2\mathbf{j} + \mathbf{k}$, $\mathbf{w} = -3\mathbf{i} + 6\mathbf{j} + 3\mathbf{k}$

8. $\mathbf{v} = \mathbf{i}$, $\mathbf{w} = 3\mathbf{i} + 2\mathbf{j} + 4\mathbf{k}$

9. Let $\mathbf{v} = (8,4,3)$ and $\mathbf{w} = (-2,1,4)$. Is $\mathbf{v} \perp \mathbf{w}$? Justify your answer.

10. Let $\mathbf{v} = (6,0,4)$ and $\mathbf{w} = (0,2,-1)$. Is $\mathbf{v} \perp \mathbf{w}$? Justify your answer.

11. For $\mathbf{v}$, $\mathbf{w}$ from Exercise 5, verify the Cauchy-Schwarz Inequality $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \, \|\mathbf{w}\|$.

12. For $\mathbf{v}$, $\mathbf{w}$ from Exercise 6, verify the Cauchy-Schwarz Inequality $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \, \|\mathbf{w}\|$.

13. For $\mathbf{v}$, $\mathbf{w}$ from Exercise 5, verify the Triangle Inequality $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.

14. For $\mathbf{v}$, $\mathbf{w}$ from Exercise 6, verify the Triangle Inequality $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.

B

Note: Consider only vectors in $\mathbb{R}^3$ for Exercises 15-25.

15. Prove Theorem 1.9(a).

16. Prove Theorem 1.9(b).

17. Prove Theorem 1.9(c).

18. Prove Theorem 1.9(d).

19. Prove Theorem 1.9(e).

20. Prove Theorem 1.10(a).

21. Prove or give a counterexample: If $\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{w}$, then $\mathbf{v} = \mathbf{w}$.

C

22. Prove or give a counterexample: If $\mathbf{v} \cdot \mathbf{w} = 0$ for all $\mathbf{v}$, then $\mathbf{w} = \mathbf{0}$.

23. Prove or give a counterexample: If $\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{w}$ for all $\mathbf{u}$, then $\mathbf{v} = \mathbf{w}$.

24. Prove that $\big|\|\mathbf{v}\| - \|\mathbf{w}\|\big| \leq \|\mathbf{v} - \mathbf{w}\|$ for all $\mathbf{v}$, $\mathbf{w}$.

25. For nonzero vectors $\mathbf{v}$ and $\mathbf{w}$, the *projection* of $\mathbf{v}$ onto $\mathbf{w}$ (sometimes written as $proj_{\mathbf{w}}\mathbf{v}$) is the vector $\mathbf{u}$ along the same line L as $\mathbf{w}$ whose terminal point is obtained by dropping a perpendicular line from the terminal point of $\mathbf{v}$ to L (see Figure 1.3.5). Show that

$$\|\mathbf{u}\| = \frac{|\mathbf{v} \cdot \mathbf{w}|}{\|\mathbf{w}\|}.$$

(Hint: Consider the angle between $\mathbf{v}$ *and* $\mathbf{w}$*.)*

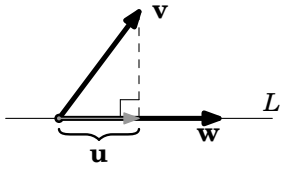


Figure 1.3.5

26. Let α, β, and γ be the angles between a nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$ and the vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, respectively. Show that $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$.
(Note: α, β, γ are often called the *direction angles* of $\mathbf{v}$, and $\cos\alpha$, $\cos\beta$, $\cos\gamma$ are called the *direction cosines*.)

1.4 Cross Product

In Section 1.3 we defined the dot product, which gave a way of multiplying two vectors. The resulting product, however, was a scalar, not a vector. In this section we will define a product of two vectors that does result in another vector. This product, called the *cross product*, is only defined for vectors in $\mathbb{R}^3$. The definition may appear strange and lacking motivation, but we will see the geometric basis for it shortly.

Definition 1.8. Let $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ be vectors in $\mathbb{R}^3$. The **cross product** of $\mathbf{v}$ and $\mathbf{w}$, denoted by $\mathbf{v} \times \mathbf{w}$, is the vector in $\mathbb{R}^3$ given by:

$$\mathbf{v} \times \mathbf{w} = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) \tag{1.10}$$

Example 1.7. Find $\mathbf{i} \times \mathbf{j}$.

Solution: Since $\mathbf{i} = (1,0,0)$ and $\mathbf{j} = (0,1,0)$, then

$$\begin{aligned}\mathbf{i} \times \mathbf{j} &= ((0)(0) - (0)(1), (0)(0) - (1)(0), (1)(1) - (0)(0)) \\ &= (0,0,1) \\ &= \mathbf{k}\end{aligned}$$

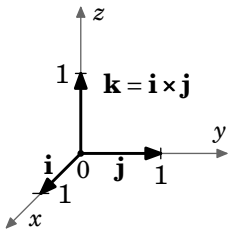


Figure 1.4.1

Similarly it can be shown that $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ and $\mathbf{k} \times \mathbf{i} = \mathbf{j}$.

In the above example, the cross product of the given vectors was perpendicular to both those vectors. It turns out that this will always be the case.

Theorem 1.11. If the cross product $\mathbf{v} \times \mathbf{w}$ of two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ is also a nonzero vector, then it is perpendicular to both $\mathbf{v}$ and $\mathbf{w}$.

Proof: We will show that $(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{v} = 0$:

$$\begin{aligned}(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{v} &= (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) \cdot (v_1, v_2, v_3) \\ &= v_2w_3v_1 - v_3w_2v_1 + v_3w_1v_2 - v_1w_3v_2 + v_1w_2v_3 - v_2w_1v_3 \\ &= v_1v_2w_3 - v_1v_2w_3 + w_1v_2v_3 - w_1v_2v_3 + v_1w_2v_3 - v_1w_2v_3 \\ &= 0 \text{ , after rearranging the terms.}\end{aligned}$$

$\therefore \mathbf{v} \times \mathbf{w} \perp \mathbf{v}$ by Corollary 1.7.
The proof that $\mathbf{v} \times \mathbf{w} \perp \mathbf{w}$ is similar. **QED**

As a consequence of the above theorem and Theorem 1.9, we have the following:

Corollary 1.12. If the cross product $\mathbf{v} \times \mathbf{w}$ of two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ is also a nonzero vector, then it is perpendicular to the span of $\mathbf{v}$ and $\mathbf{w}$.

The span of any two nonzero, nonparallel vectors $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$ is a plane P, so the above corollary shows that $\mathbf{v} \times \mathbf{w}$ is perpendicular to that plane. As shown in Figure 1.4.2, there are two possible directions for $\mathbf{v} \times \mathbf{w}$, one the opposite of the other. It turns out (see Appendix B) that the direction of $\mathbf{v} \times \mathbf{w}$ is given by the *right-hand rule*, that is, the vectors $\mathbf{v}$, $\mathbf{w}$, $\mathbf{v} \times \mathbf{w}$ form a right-handed system. Recall from Section 1.1 that this means that you can point your thumb upwards in the direction of $\mathbf{v} \times \mathbf{w}$ while rotating $\mathbf{v}$ towards $\mathbf{w}$ with the remaining four fingers.

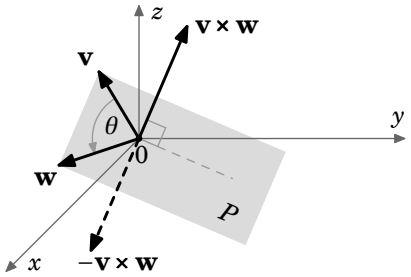


Figure 1.4.2 Direction of $\mathbf{v} \times \mathbf{w}$

We will now derive a formula for the magnitude of $\mathbf{v} \times \mathbf{w}$, for nonzero vectors $\mathbf{v}$, $\mathbf{w}$:

$$\begin{aligned}
\|\mathbf{v} \times \mathbf{w}\|^2 &= (v_2w_3 - v_3w_2)^2 + (v_3w_1 - v_1w_3)^2 + (v_1w_2 - v_2w_1)^2 \\
&= v_2^2w_3^2 - 2v_2w_2v_3w_3 + v_3^2w_2^2 + v_3^2w_1^2 - 2v_1w_1v_3w_3 + v_1^2w_3^2 + v_1^2w_2^2 - 2v_1w_1v_2w_2 + v_2^2w_1^2 \\
&= v_1^2(w_2^2 + w_3^2) + v_2^2(w_1^2 + w_3^2) + v_3^2(w_1^2 + w_2^2) - 2(v_1w_1v_2w_2 + v_1w_1v_3w_3 + v_2w_2v_3w_3)
\end{aligned}$$

and now adding and subtracting $v_1^2w_1^2$, $v_2^2w_2^2$, and $v_3^2w_3^2$ on the right side gives

$$\begin{aligned}
&= v_1^2(w_1^2 + w_2^2 + w_3^2) + v_2^2(w_1^2 + w_2^2 + w_3^2) + v_3^2(w_1^2 + w_2^2 + w_3^2) \\
&\quad - (v_1^2w_1^2 + v_2^2w_2^2 + v_3^2w_3^2 + 2(v_1w_1v_2w_2 + v_1w_1v_3w_3 + v_2w_2v_3w_3)) \\
&= (v_1^2 + v_2^2 + v_3^2)(w_1^2 + w_2^2 + w_3^2) \\
&\quad - ((v_1w_1)^2 + (v_2w_2)^2 + (v_3w_3)^2 + 2(v_1w_1)(v_2w_2) + 2(v_1w_1)(v_3w_3) + 2(v_2w_2)(v_3w_3))
\end{aligned}$$

so using $(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$ for the subtracted term gives

$$\begin{aligned}
&= (v_1^2 + v_2^2 + v_3^2)(w_1^2 + w_2^2 + w_3^2) - (v_1w_1 + v_2w_2 + v_3w_3)^2 \\
&= \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 - (\mathbf{v} \cdot \mathbf{w})^2 \\
&= \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 \left(1 - \frac{(\mathbf{v} \cdot \mathbf{w})^2}{\|\mathbf{v}\|^2 \|\mathbf{w}\|^2}\right), \text{ since } \|\mathbf{v}\| > 0 \text{ and } \|\mathbf{w}\| > 0, \text{ so by Theorem 1.6} \\
&= \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 (1 - \cos^2\theta) \text{ , where } \theta \text{ is the angle between } \mathbf{v} \text{ and } \mathbf{w}, \text{ so}
\end{aligned}$$

$\|\mathbf{v} \times \mathbf{w}\|^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 \sin^2\theta$, and since $0° \le \theta \le 180°$, then $\sin\theta \ge 0$, so we have:

If θ is the angle between nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$, then

$$\|\mathbf{v} \times \mathbf{w}\| = \|\mathbf{v}\| \, \|\mathbf{w}\| \sin\theta \tag{1.11}$$

It may seem strange to bother with the above formula, when the magnitude of the cross product can be calculated directly, like for any other vector. The formula is more useful for its applications in geometry, as in the following example.

Example 1.8. Let $\triangle PQR$ and $PQRS$ be a triangle and parallelogram, respectively, as shown in Figure 1.4.3.

Figure 1.4.3

Think of the triangle as existing in $\mathbb{R}^3$, and identify the sides QR and QP with vectors $\mathbf{v}$ and $\mathbf{w}$, respectively, in $\mathbb{R}^3$. Let θ be the angle between $\mathbf{v}$ and $\mathbf{w}$. The area A_{PQR} of $\triangle PQR$ is $\frac{1}{2}bh$, where b is the base of the triangle and h is the height. So we see that

$$b = \|\mathbf{v}\| \quad \text{and} \quad h = \|\mathbf{w}\| \sin\theta$$

$$\begin{aligned} A_{PQR} &= \frac{1}{2} \|\mathbf{v}\| \, \|\mathbf{w}\| \sin\theta \\ &= \frac{1}{2} \|\mathbf{v} \times \mathbf{w}\| \end{aligned}$$

So since the area A_{PQRS} of the parallelogram $PQRS$ is twice the area of the triangle $\triangle PQR$, then

$$A_{PQRS} = \|\mathbf{v}\| \, \|\mathbf{w}\| \sin\theta$$

By the discussion in Example 1.8, we have proved the following theorem:

Theorem 1.13. Area of triangles and parallelograms

(a) The area A of a triangle with adjacent sides $\mathbf{v}$, $\mathbf{w}$ (as vectors in $\mathbb{R}^3$) is:

$$A = \frac{1}{2} \|\mathbf{v} \times \mathbf{w}\|$$

(b) The area A of a parallelogram with adjacent sides $\mathbf{v}$, $\mathbf{w}$ (as vectors in $\mathbb{R}^3$) is:

$$A = \|\mathbf{v} \times \mathbf{w}\|$$

It may seem at first glance that since the formulas derived in Example 1.8 were for the adjacent sides QP and QR only, then the more general statements in Theorem 1.13 that the formulas hold for *any* adjacent sides are not justified. We would get a different *formula* for the area if we had picked PQ and PR as the adjacent sides, but it can be shown (see Exercise 26) that the different formulas would yield the same value, so the choice of adjacent sides indeed does not matter, and Theorem 1.13 is valid.

Theorem 1.13 makes it simpler to calculate the area of a triangle in 3-dimensional space than by using traditional geometric methods.

Example 1.9. Calculate the area of the triangle $\triangle PQR$, where $P = (2,4,-7)$, $Q = (3,7,18)$, and $R = (-5,12,8)$.

Solution: Let $\mathbf{v} = \overrightarrow{PQ}$ and $\mathbf{w} = \overrightarrow{PR}$, as in Figure 1.4.4. Then $\mathbf{v} = (3,7,18)-(2,4,-7) = (1,3,25)$ and $\mathbf{w} = (-5,12,8)-(2,4,-7) = (-7,8,15)$, so the area A of the triangle $\triangle PQR$ is

$$\begin{aligned} A &= \frac{1}{2}\|\mathbf{v}\times\mathbf{w}\| = \frac{1}{2}\|(1,3,25)\times(-7,8,15)\| \\ &= \frac{1}{2}\big\|((3)(15)-(25)(8),(25)(-7)-(1)(15),(1)(8)-(3)(-7))\big\| \\ &= \frac{1}{2}\big\|(-155,-190,29)\big\| \\ &= \frac{1}{2}\sqrt{(-155)^2+(-190)^2+29^2} = \frac{1}{2}\sqrt{60966} \\ A &\approx 123.46 \end{aligned}$$

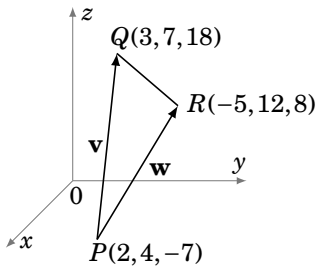


Figure 1.4.4

Example 1.10. Calculate the area of the parallelogram $PQRS$, where $P = (1,1)$, $Q = (2,3)$, $R = (5,4)$, and $S = (4,2)$.

Solution: Let $\mathbf{v} = \overrightarrow{SP}$ and $\mathbf{w} = \overrightarrow{SR}$, as in Figure 1.4.5. Then $\mathbf{v} = (1,1)-(4,2) = (-3,-1)$ and $\mathbf{w} = (5,4)-(4,2) = (1,2)$. But these are vectors in $\mathbb{R}^2$, and the cross product is only defined for vectors in $\mathbb{R}^3$. However, $\mathbb{R}^2$ can be thought of as the subset of $\mathbb{R}^3$ such that the z-coordinate is always 0. So we can write $\mathbf{v} = (-3,-1,0)$ and $\mathbf{w} = (1,2,0)$. Then the area A of $PQRS$ is

$$\begin{aligned} A &= \|\mathbf{v}\times\mathbf{w}\| = \big\|(-3,-1,0)\times(1,2,0)\big\| \\ &= \big\|((-1)(0)-(0)(2),(0)(1)-(-3)(0),(-3)(2)-(-1)(1))\big\| \\ &= \big\|(0,0,-5)\big\| \\ A &= 5 \end{aligned}$$

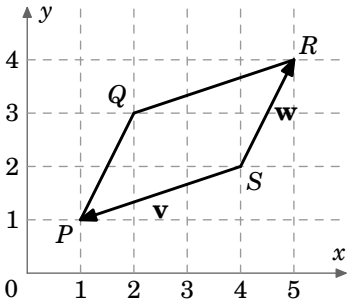


Figure 1.4.5

The following theorem summarizes the basic properties of the cross product.

Theorem 1.14. For any vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$, and scalar k, we have

(a) $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$	Anticommutative Law
(b) $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$	Distributive Law
(c) $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$	Distributive Law
(d) $(k\mathbf{v}) \times \mathbf{w} = \mathbf{v} \times (k\mathbf{w}) = k(\mathbf{v} \times \mathbf{w})$	Associative Law
(e) $\mathbf{v} \times \mathbf{0} = \mathbf{0} = \mathbf{0} \times \mathbf{v}$	
(f) $\mathbf{v} \times \mathbf{v} = \mathbf{0}$	
(g) $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ if and only if $\mathbf{v} \parallel \mathbf{w}$	

Proof: The proofs of properties (b)-(f) are straightforward. We will prove parts (a) and (g) and leave the rest to the reader as exercises.

(a) By the definition of the cross product and scalar multiplication, we have:

$$\begin{aligned}\mathbf{v} \times \mathbf{w} &= (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) \\ &= -(v_3w_2 - v_2w_3, v_1w_3 - v_3w_1, v_2w_1 - v_1w_2) \\ &= -(w_2v_3 - w_3v_2, w_3v_1 - w_1v_3, w_1v_2 - w_2v_1) \\ &= -\mathbf{w} \times \mathbf{v}\end{aligned}$$

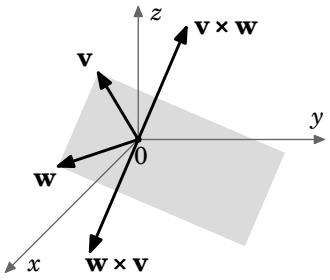


Figure 1.4.6

Note that this says that $\mathbf{v} \times \mathbf{w}$ and $\mathbf{w} \times \mathbf{v}$ have the same magnitude but opposite direction (see Figure 1.4.6).

(g) If either $\mathbf{v}$ or $\mathbf{w}$ is $\mathbf{0}$ then $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ by part (e), and either $\mathbf{v} = \mathbf{0} = 0\mathbf{w}$ or $\mathbf{w} = \mathbf{0} = 0\mathbf{v}$, so $\mathbf{v}$ and $\mathbf{w}$ are scalar multiples, i.e. they are parallel.

If both $\mathbf{v}$ and $\mathbf{w}$ are nonzero, and θ is the angle between them, then by formula (1.11), $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ if and only if $\|\mathbf{v}\| \|\mathbf{w}\| \sin\theta = 0$, which is true if and only if $\sin\theta = 0$ (since $\|\mathbf{v}\| > 0$ and $\|\mathbf{w}\| > 0$). So since $0° \le \theta \le 180°$, then $\sin\theta = 0$ if and only if $\theta = 0°$ or $180°$. But the angle between $\mathbf{v}$ and $\mathbf{w}$ is $0°$ or $180°$ if and only if $\mathbf{v} \parallel \mathbf{w}$. **QED**

Example 1.11. Adding to Example 1.7, we have

$$\mathbf{i} \times \mathbf{j} = \mathbf{k} \qquad \mathbf{j} \times \mathbf{k} = \mathbf{i} \qquad \mathbf{k} \times \mathbf{i} = \mathbf{j}$$
$$\mathbf{j} \times \mathbf{i} = -\mathbf{k} \qquad \mathbf{k} \times \mathbf{j} = -\mathbf{i} \qquad \mathbf{i} \times \mathbf{k} = -\mathbf{j}$$
$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}$$

Recall from geometry that a *parallelepiped* is a 3-dimensional solid with 6 faces, all of which are parallelograms.[6]

[6] An equivalent definition of a parallelepiped is: the collection of all scalar combinations $k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + k_3\mathbf{v}_3$ of some vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ in $\mathbb{R}^3$, where $0 \le k_1, k_2, k_3 \le 1$.

Example 1.12. *Volume of a parallelepiped:* Let the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$ represent adjacent sides of a parallelepiped P, with $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ forming a right-handed system, as in Figure 1.4.7. Show that the volume of P is the *scalar triple product* $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$.

Solution: Recall that the volume vol(P) of a parallelepiped P is the area A of the base parallelogram times the height h. By Theorem 1.13(b), the area A of the base parallelogram is $\|\mathbf{v}\times\mathbf{w}\|$. And we can see that since $\mathbf{v}\times\mathbf{w}$ is perpendicular to the base parallelogram determined by $\mathbf{v}$ and $\mathbf{w}$, then the height h is $\|\mathbf{u}\|\cos\theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}\times\mathbf{w}$. By Theorem 1.6 we know that

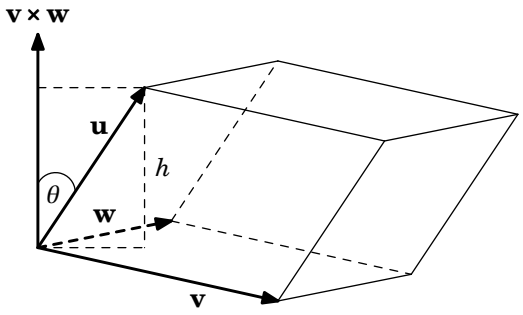


Figure 1.4.7 Parallelepiped P

$$\cos\theta = \frac{\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})}{\|\mathbf{u}\|\,\|\mathbf{v}\times\mathbf{w}\|}. \text{ Hence,}$$

$$\begin{aligned}\text{vol}(P) &= A\,h \\ &= \|\mathbf{v}\times\mathbf{w}\|\,\frac{\|\mathbf{u}\|\,\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})}{\|\mathbf{u}\|\,\|\mathbf{v}\times\mathbf{w}\|} \\ &= \mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})\end{aligned}$$

In Example 1.12 the height h of the parallelepiped is $\|\mathbf{u}\|\cos\theta$, and not $-\|\mathbf{u}\|\cos\theta$, because the vector $\mathbf{u}$ is on the same side of the base parallelogram's plane as the vector $\mathbf{v}\times\mathbf{w}$ (so that $\cos\theta > 0$). Since the volume is the same no matter which base and height we use, then repeating the same steps using the base determined by $\mathbf{u}$ and $\mathbf{v}$ (since $\mathbf{w}$ is on the same side of that base's plane as $\mathbf{u}\times\mathbf{v}$), the volume is $\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v})$. Repeating this with the base determined by $\mathbf{w}$ and $\mathbf{u}$, we have the following result:

For any vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$,

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) = \mathbf{w}\cdot(\mathbf{u}\times\mathbf{v}) = \mathbf{v}\cdot(\mathbf{w}\times\mathbf{u}) \tag{1.12}$$

(Note that the equalities hold trivially if any of the vectors are $\mathbf{0}$.)

Since $\mathbf{v}\times\mathbf{w} = -\mathbf{w}\times\mathbf{v}$ for any vectors $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$, then picking the wrong order for the three adjacent sides in the scalar triple product in formula (1.12) will give you the negative of the volume of the parallelepiped. So taking the absolute value of the scalar triple product for *any* order of the three adjacent sides will *always* give the volume:

Theorem 1.15. If vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$ represent any three adjacent sides of a parallelepiped, then the volume of the parallelepiped is $|\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})|$.

Another type of triple product is the *vector triple product* $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})$. The proof of the following theorem is left as an exercise for the reader:

Theorem 1.16. For any vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$,

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} \tag{1.13}$$

An examination of the formula in Theorem 1.16 gives some idea of the geometry of the vector triple product. By the right side of formula (1.13), we see that $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ is a scalar combination of $\mathbf{v}$ and $\mathbf{w}$, and hence lies in the plane containing $\mathbf{v}$ and $\mathbf{w}$ (i.e. $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$, $\mathbf{v}$ and $\mathbf{w}$ are **coplanar**). This makes sense since, by Theorem 1.11, $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ is perpendicular to both $\mathbf{u}$ and $\mathbf{v} \times \mathbf{w}$. In particular, being perpendicular to $\mathbf{v} \times \mathbf{w}$ means that $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ lies in the plane containing $\mathbf{v}$ and $\mathbf{w}$, since that plane is itself perpendicular to $\mathbf{v} \times \mathbf{w}$. But then how is $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ also perpendicular to $\mathbf{u}$, which could be any vector? The following example may help to see how this works.

Example 1.13. Find $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ for $\mathbf{u} = (1,2,4)$, $\mathbf{v} = (2,2,0)$, $\mathbf{w} = (1,3,0)$.

Solution: Since $\mathbf{u} \cdot \mathbf{v} = 6$ and $\mathbf{u} \cdot \mathbf{w} = 7$, then

$$\begin{aligned} \mathbf{u} \times (\mathbf{v} \times \mathbf{w}) &= (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} \\ &= 7(2,2,0) - 6(1,3,0) = (14,14,0) - (6,18,0) \\ &= (8,-4,0) \end{aligned}$$

Note that $\mathbf{v}$ and $\mathbf{w}$ lie in the xy-plane, and that $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ also lies in that plane. Also, $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ is perpendicular to both $\mathbf{u}$ and $\mathbf{v} \times \mathbf{w} = (0,0,4)$ (see Figure 1.4.8).

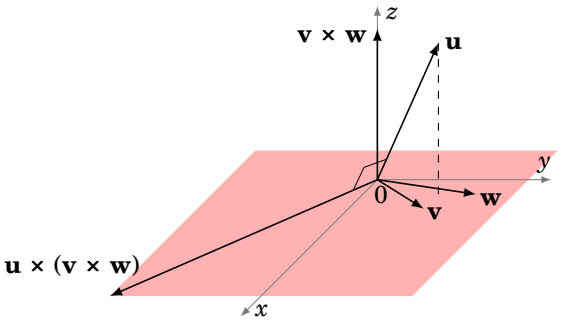


Figure 1.4.8

For vectors $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ and $\mathbf{w} = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}$ in component form, the cross product is written as: $\mathbf{v} \times \mathbf{w} = (v_2w_3 - v_3w_2)\mathbf{i} + (v_3w_1 - v_1w_3)\mathbf{j} + (v_1w_2 - v_2w_1)\mathbf{k}$. It is often easier to use the component form for the cross product, because it can be represented as a *determinant*. We will not go too deeply into the theory of determinants[7]; we will just cover what is essential for our purposes.

[7] See ANTON and RORRES for a fuller development.

A **2 × 2 matrix** is an array of two rows and two columns of scalars, written as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{or} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where a, b, c, d are scalars. The **determinant** of such a matrix, written as

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad \text{or} \quad \det \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

is the scalar defined by the following formula:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

It may help to remember this formula as being the product of the scalars on the downward diagonal minus the product of the scalars on the upward diagonal.

Example 1.14.

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = (1)(4) - (2)(3) = 4 - 6 = -2$$

A **3 × 3 matrix** is an array of three rows and three columns of scalars, written as

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \quad \text{or} \quad \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix},$$

and its determinant is given by the formula:

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \tag{1.14}$$

One way to remember the above formula is the following: multiply each scalar in the first row by the determinant of the 2×2 matrix that remains after removing the row and column that contain that scalar, then sum those products up, putting alternating plus and minus signs in front of each (starting with a plus).

Example 1.15.

$$\begin{vmatrix} 1 & 0 & 2 \\ 4 & -1 & 3 \\ 1 & 0 & 2 \end{vmatrix} = 1 \begin{vmatrix} -1 & 3 \\ 0 & 2 \end{vmatrix} - 0 \begin{vmatrix} 4 & 3 \\ 1 & 2 \end{vmatrix} + 2 \begin{vmatrix} 4 & -1 \\ 1 & 0 \end{vmatrix} = 1(-2-0) - 0(8-3) + 2(0+1) = 0$$

We defined the determinant as a scalar, derived from algebraic operations on scalar entries in a matrix. However, if we put three *vectors* in the first row of a 3×3 matrix, then the definition still makes sense, since we would be performing scalar multiplication on those three vectors (they would be multiplied by the 2×2 scalar determinants as before). This gives us a determinant that is now a vector, and lets us write the cross product of $\mathbf{v}=v_1\mathbf{i}+v_2\mathbf{j}+v_3\mathbf{k}$ and $\mathbf{w}=w_1\mathbf{i}+w_2\mathbf{j}+w_3\mathbf{k}$ as a determinant:

$$\mathbf{v}\times\mathbf{w}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\ v_1&v_2&v_3\\ w_1&w_2&w_3\end{vmatrix}=\begin{vmatrix}v_2&v_3\\ w_2&w_3\end{vmatrix}\mathbf{i}-\begin{vmatrix}v_1&v_3\\ w_1&w_3\end{vmatrix}\mathbf{j}+\begin{vmatrix}v_1&v_2\\ w_1&w_2\end{vmatrix}\mathbf{k}$$
$$=(v_2w_3-v_3w_2)\mathbf{i}+(v_3w_1-v_1w_3)\mathbf{j}+(v_1w_2-v_2w_1)\mathbf{k}$$

Example 1.16. Let $\mathbf{v}=4\mathbf{i}-\mathbf{j}+3\mathbf{k}$ and $\mathbf{w}=\mathbf{i}+2\mathbf{k}$. Then

$$\mathbf{v}\times\mathbf{w}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\ 4&-1&3\\ 1&0&2\end{vmatrix}=\begin{vmatrix}-1&3\\ 0&2\end{vmatrix}\mathbf{i}-\begin{vmatrix}4&3\\ 1&2\end{vmatrix}\mathbf{j}+\begin{vmatrix}4&-1\\ 1&0\end{vmatrix}\mathbf{k}=-2\mathbf{i}-5\mathbf{j}+\mathbf{k}$$

The scalar triple product can also be written as a determinant. In fact, by Example 1.12, the following theorem provides an alternate definition of the determinant of a 3×3 matrix as the volume of a parallelepiped whose adjacent sides are the rows of the matrix and form a right-handed system (a left-handed system would give the negative volume).

Theorem 1.17. For any vectors $\mathbf{u}=(u_1,u_2,u_3)$, $\mathbf{v}=(v_1,v_2,v_3)$, $\mathbf{w}=(w_1,w_2,w_3)$ in $\mathbb{R}^3$:

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})=\begin{vmatrix}u_1&u_2&u_3\\ v_1&v_2&v_3\\ w_1&w_2&w_3\end{vmatrix}\tag{1.15}$$

Example 1.17. Find the volume of the parallelepiped with adjacent sides $\mathbf{u}=(2,1,3)$, $\mathbf{v}=(-1,3,2)$, $\mathbf{w}=(1,1,-2)$ (see Figure 1.4.9).

Solution: By Theorem 1.15, the volume $\text{vol}(P)$ of the parallelepiped P is the absolute value of the scalar triple product of the three adjacent sides (in any order). By Theorem 1.17,

$$\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})=\begin{vmatrix}2&1&3\\ -1&3&2\\ 1&1&-2\end{vmatrix}$$
$$=2\begin{vmatrix}3&2\\ 1&-2\end{vmatrix}-1\begin{vmatrix}-1&2\\ 1&-2\end{vmatrix}+3\begin{vmatrix}-1&3\\ 1&1\end{vmatrix}$$
$$=2(-8)-1(0)+3(-4)=-28,\text{ so}$$
$$\text{vol}(P)=|-28|=28.$$

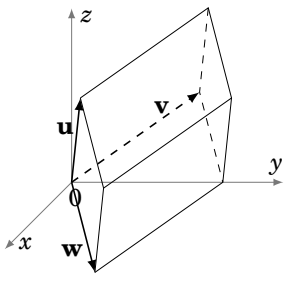


Figure 1.4.9 P

Interchanging the dot and cross products can be useful in proving vector identities:

Example 1.18. Prove: $(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{w}\times\mathbf{z}) = \begin{vmatrix}\mathbf{u}\cdot\mathbf{w} & \mathbf{u}\cdot\mathbf{z}\\ \mathbf{v}\cdot\mathbf{w} & \mathbf{v}\cdot\mathbf{z}\end{vmatrix}$ for all vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{z}$ in $\mathbb{R}^3$.

Solution: Let $\mathbf{x} = \mathbf{u}\times\mathbf{v}$. Then

$$\begin{aligned}(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{w}\times\mathbf{z}) &= \mathbf{x}\cdot(\mathbf{w}\times\mathbf{z})\\ &= \mathbf{w}\cdot(\mathbf{z}\times\mathbf{x}) \quad \text{(by formula (1.12))}\\ &= \mathbf{w}\cdot(\mathbf{z}\times(\mathbf{u}\times\mathbf{v}))\\ &= \mathbf{w}\cdot((\mathbf{z}\cdot\mathbf{v})\mathbf{u}-(\mathbf{z}\cdot\mathbf{u})\mathbf{v}) \quad \text{(by Theorem 1.16)}\\ &= (\mathbf{z}\cdot\mathbf{v})(\mathbf{w}\cdot\mathbf{u})-(\mathbf{z}\cdot\mathbf{u})(\mathbf{w}\cdot\mathbf{v})\\ &= (\mathbf{u}\cdot\mathbf{w})(\mathbf{v}\cdot\mathbf{z})-(\mathbf{u}\cdot\mathbf{z})(\mathbf{v}\cdot\mathbf{w}) \quad \text{(by commutativity of the dot product).}\\ &= \begin{vmatrix}\mathbf{u}\cdot\mathbf{w} & \mathbf{u}\cdot\mathbf{z}\\ \mathbf{v}\cdot\mathbf{w} & \mathbf{v}\cdot\mathbf{z}\end{vmatrix}\end{aligned}$$

Exercises

A

For Exercises 1-6, calculate $\mathbf{v}\times\mathbf{w}$.

1. $\mathbf{v} = (5,1,-2)$, $\mathbf{w} = (4,-4,3)$

2. $\mathbf{v} = (7,2,-10)$, $\mathbf{w} = (2,6,4)$

3. $\mathbf{v} = (2,1,4)$, $\mathbf{w} = (1,-2,0)$

4. $\mathbf{v} = (1,3,2)$, $\mathbf{w} = (7,2,-10)$

5. $\mathbf{v} = -\mathbf{i}+2\mathbf{j}+\mathbf{k}$, $\mathbf{w} = -3\mathbf{i}+6\mathbf{j}+3\mathbf{k}$

6. $\mathbf{v} = \mathbf{i}$, $\mathbf{w} = 3\mathbf{i}+2\mathbf{j}+4\mathbf{k}$

For Exercises 7-8, calculate the area of the triangle $\triangle PQR$.

7. $P = (5,1,-2)$, $Q = (4,-4,3)$, $R = (2,4,0)$

8. $P = (4,0,2)$, $Q = (2,1,5)$, $R = (-1,0,-1)$

For Exercises 9-10, calculate the area of the parallelogram $PQRS$.

9. $P = (2,1,3)$, $Q = (1,4,5)$, $R = (2,5,3)$, $S = (3,2,1)$

10. $P = (-2,-2)$, $Q = (1,4)$, $R = (6,6)$, $S = (3,0)$

For Exercises 11-12, find the volume of the parallelepiped with adjacent sides $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$.

11. $\mathbf{u} = (1,1,3)$, $\mathbf{v} = (2,1,4)$, $\mathbf{w} = (5,1,-2)$

12. $\mathbf{u} = (1,3,2)$, $\mathbf{v} = (7,2,-10)$, $\mathbf{w} = (1,0,1)$

For Exercises 13-14, calculate $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ and $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})$.

13. $\mathbf{u} = (1,1,1)$, $\mathbf{v} = (3,0,2)$, $\mathbf{w} = (2,2,2)$

14. $\mathbf{u} = (1,0,2)$, $\mathbf{v} = (-1,0,3)$, $\mathbf{w} = (2,0,-2)$

15. Calculate $(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{w}\times\mathbf{z})$ for $\mathbf{u} = (1,1,1)$, $\mathbf{v} = (3,0,2)$, $\mathbf{w} = (2,2,2)$, $\mathbf{z} = (2,1,4)$.

B

16. If $\mathbf{v}$ and $\mathbf{w}$ are unit vectors in $\mathbb{R}^3$, under what condition(s) would $\mathbf{v} \times \mathbf{w}$ also be a unit vector in $\mathbb{R}^3$? Justify your answer.

17. Show that if $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ for all $\mathbf{w}$ in $\mathbb{R}^3$, then $\mathbf{v} = \mathbf{0}$.

18. Prove Theorem 1.14(b).

19. Prove Theorem 1.14(c).

20. Prove Theorem 1.14(d).

21. Prove Theorem 1.14(e).

22. Prove Theorem 1.14(f).

23. Prove Theorem 1.16.

24. Prove Theorem 1.17. (*Hint: Expand both sides of the equation.*)

25. Prove the following for all vectors $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^3$:

(a) $\|\mathbf{v} \times \mathbf{w}\|^2 + |\mathbf{v} \cdot \mathbf{w}|^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$

(b) If $\mathbf{v} \cdot \mathbf{w} = 0$ and $\mathbf{v} \times \mathbf{w} = \mathbf{0}$, then $\mathbf{v} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$.

C

26. Prove that in Example 1.8 the formula for the area of the triangle $\triangle PQR$ yields the same value no matter which two adjacent sides are chosen. To do this, show that $\frac{1}{2}\|\mathbf{u} \times (-\mathbf{w})\| = \frac{1}{2}\|\mathbf{v} \times \mathbf{w}\|$, where $\mathbf{u} = PR$, $-\mathbf{w} = PQ$, and $\mathbf{v} = QR$, $\mathbf{w} = QP$ as before. Similarly, show that $\frac{1}{2}\|(-\mathbf{u}) \times (-\mathbf{v})\| = \frac{1}{2}\|\mathbf{v} \times \mathbf{w}\|$, where $-\mathbf{u} = RP$ and $-\mathbf{v} = RQ$.

27. Consider the vector equation $\mathbf{a} \times \mathbf{x} = \mathbf{b}$ in $\mathbb{R}^3$, where $\mathbf{a} \neq \mathbf{0}$. Show that:

(a) $\mathbf{a} \cdot \mathbf{b} = 0$

(b) $\mathbf{x} = \dfrac{\mathbf{b} \times \mathbf{a}}{\|\mathbf{a}\|^2} + k\mathbf{a}$ is a solution to the equation, for any scalar k

28. Prove the *Jacobi identity*: $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) + \mathbf{v} \times (\mathbf{w} \times \mathbf{u}) + \mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = \mathbf{0}$

29. Show that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ lie in the same plane in $\mathbb{R}^3$ if and only if $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = 0$.

30. For all vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{z}$ in $\mathbb{R}^3$, show that

$$(\mathbf{u} \times \mathbf{v}) \times (\mathbf{w} \times \mathbf{z}) = (\mathbf{z} \cdot (\mathbf{u} \times \mathbf{v}))\mathbf{w} - (\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}))\mathbf{z}$$

and that

$$(\mathbf{u} \times \mathbf{v}) \times (\mathbf{w} \times \mathbf{z}) = (\mathbf{u} \cdot (\mathbf{w} \times \mathbf{z}))\mathbf{v} - (\mathbf{v} \cdot (\mathbf{w} \times \mathbf{z}))\mathbf{u}$$

Why do both equations make sense geometrically?

1.5 Lines and Planes

Now that we know how to perform some operations on vectors, we can start to deal with some familiar geometric objects, like lines and planes, in the language of vectors. The reason for doing this is simple: using vectors makes it easier to study objects in 3-dimensional Euclidean space. We will first consider lines.

Line through a point, parallel to a vector

Let $P = (x_0, y_0, z_0)$ be a point in $\mathbb{R}^3$, let $\mathbf{v} = (a, b, c)$ be a nonzero vector, and let L be the line through P which is parallel to $\mathbf{v}$ (see Figure 1.5.1).

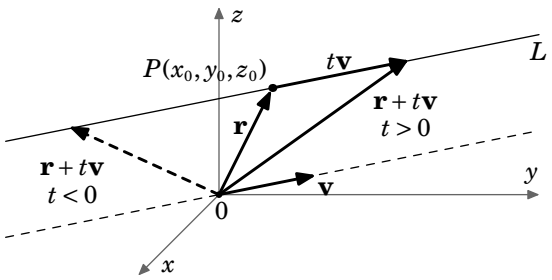


Figure 1.5.1

Let $\mathbf{r} = (x_0, y_0, z_0)$ be the *vector* pointing from the origin to P. Since multiplying the vector $\mathbf{v}$ by a scalar t lengthens or shrinks $\mathbf{v}$ while preserving its direction if $t > 0$, and reversing its direction if $t < 0$, then we see from Figure 1.5.1 that every point on the line L can be obtained by adding the vector $t\mathbf{v}$ to the vector $\mathbf{r}$ for some scalar t. That is, as t varies over all real numbers, the vector $\mathbf{r} + t\mathbf{v}$ will point to every point on L. We can summarize the *vector representation of* L as follows:

For a point $P = (x_0, y_0, z_0)$ and nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$, the line L through P parallel to $\mathbf{v}$ is given by

$$\mathbf{r} + t\mathbf{v}, \quad \text{for } -\infty < t < \infty \tag{1.16}$$

where $\mathbf{r} = (x_0, y_0, z_0)$ is the vector pointing to P.

Note that we used the correspondence between a vector and its terminal point. Since $\mathbf{v} = (a, b, c)$, then the terminal point of the vector $\mathbf{r} + t\mathbf{v}$ is $(x_0 + at, y_0 + bt, z_0 + ct)$. We then get the *parametric representation of* L with the *parameter* t:

For a point $P = (x_0, y_0, z_0)$ and nonzero vector $\mathbf{v} = (a, b, c)$ in $\mathbb{R}^3$, the line L through P parallel to $\mathbf{v}$ consists of all points (x, y, z) given by

$$x = x_0 + at, \quad y = y_0 + bt, \quad z = z_0 + ct, \quad \text{for } -\infty < t < \infty \tag{1.17}$$

Note that in both representations we get the point P on L by letting $t = 0$.

In formula (1.17), if $a \neq 0$, then we can solve for the parameter t: $t = (x - x_0)/a$. We can also solve for t in terms of y and in terms of z if neither b nor c, respectively, is zero: $t = (y - y_0)/b$ and $t = (z - z_0)/c$. These three values all equal the same value t, so we can write the following system of equalities, called the *symmetric representation of* L:

For a point $P = (x_0, y_0, z_0)$ and vector $\mathbf{v} = (a, b, c)$ in $\mathbb{R}^3$ with a, b and c all nonzero, the line L through P parallel to $\mathbf{v}$ consists of all points (x, y, z) given by the equations

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c} \tag{1.18}$$

What if, say, $a = 0$ in the above scenario? We can not divide by zero, but we do know that $x = x_0 + at$, and so $x = x_0 + 0t = x_0$. Then the symmetric representation of L would be:

$$x = x_0, \quad \frac{y - y_0}{b} = \frac{z - z_0}{c} \tag{1.19}$$

Note that this says that the line L lies in the *plane* $x = x_0$, which is parallel to the yz-plane (see Figure 1.5.2). Similar equations can be derived for the cases when $b = 0$ or $c = 0$.

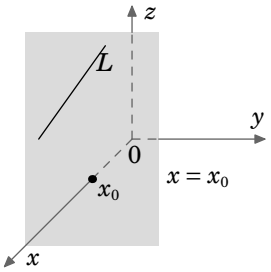


Figure 1.5.2

You may have noticed that the vector representation of L in formula (1.16) is more compact than the parametric and symmetric formulas. That is an advantage of using vector notation. Technically, though, the vector representation gives us the *vectors* whose terminal points make up the line L, not just L itself. So you have to remember to identify the vectors $\mathbf{r} + t\mathbf{v}$ with their terminal points. On the other hand, the parametric representation *always* gives just the points on L and nothing else.

Example 1.19. Write the line L through the point $P = (2, 3, 5)$ and parallel to the vector $\mathbf{v} = (4, -1, 6)$, in the following forms: (a) vector, (b) parametric, (c) symmetric. Lastly: (d) find two points on L distinct from P.

Solution: (a) Let $\mathbf{r} = (2, 3, 5)$. Then by formula (1.16), L is given by:

$$\mathbf{r} + t\mathbf{v} = (2, 3, 5) + t(4, -1, 6), \quad \text{for } -\infty < t < \infty$$

(b) L consists of the points (x, y, z) such that

$$x = 2 + 4t, \quad y = 3 - t, \quad z = 5 + 6t, \quad \text{for } -\infty < t < \infty$$

(c) L consists of the points (x, y, z) such that

$$\frac{x - 2}{4} = \frac{y - 3}{-1} = \frac{z - 5}{6}$$

(d) Letting $t = 1$ and $t = 2$ in part(b) yields the points $(6, 2, 11)$ and $(10, 1, 17)$ on L.

Line through two points

Let $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ be distinct points in $\mathbb{R}^3$, and let L be the line through P_1 and P_2. Let $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (x_2, y_2, z_2)$ be the vectors pointing to P_1 and P_2, respectively. Then as we can see from Figure 1.5.3, $\mathbf{r}_2 - \mathbf{r}_1$ is the vector from P_1 to P_2. So if we multiply the vector $\mathbf{r}_2 - \mathbf{r}_1$ by a scalar t and add it to the vector $\mathbf{r}_1$, we will get the entire line L as t varies over all real numbers. The following is a summary of the vector, parametric, and symmetric forms for the line L:

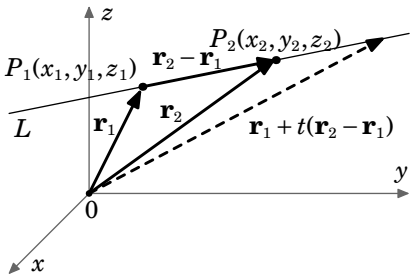


Figure 1.5.3

Let $P_1 = (x_1, y_1, z_1)$, $P_2 = (x_2, y_2, z_2)$ be distinct points in $\mathbb{R}^3$, and let $\mathbf{r}_1 = (x_1, y_1, z_1)$, $\mathbf{r}_2 = (x_2, y_2, z_2)$. Then the line L through P_1 and P_2 has the following representations:

Vector:

$$\mathbf{r}_1 + t(\mathbf{r}_2 - \mathbf{r}_1)\,, \text{ for } -\infty < t < \infty \tag{1.20}$$

Parametric:

$$x = x_1 + (x_2 - x_1)t, \quad y = y_1 + (y_2 - y_1)t, \quad z = z_1 + (z_2 - z_1)t, \text{ for } -\infty < t < \infty \tag{1.21}$$

Symmetric:

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1} \quad (\text{if } x_1 \neq x_2,\ y_1 \neq y_2, \text{ and } z_1 \neq z_2) \tag{1.22}$$

Example 1.20. Write the line L through the points $P_1 = (-3, 1, -4)$ and $P_2 = (4, 4, -6)$ in parametric form.

Solution: By formula (1.21), L consists of the points (x, y, z) such that

$$x = -3 + 7t, \quad y = 1 + 3t, \quad z = -4 - 2t, \text{ for } -\infty < t < \infty$$

Distance between a point and a line

Let L be a line in $\mathbb{R}^3$ in vector form as $\mathbf{r} + t\mathbf{v}$ (for $-\infty < t < \infty$), and let P be a point not on L. The distance d from P to L is the length of the line segment from P to L which is perpendicular to L (see Figure 1.5.4). Pick a point Q on L, and let $\mathbf{w}$ be the vector from Q to P. If θ is the angle between $\mathbf{w}$ and $\mathbf{v}$, then $d = \|\mathbf{w}\| \sin\theta$. So since $\|\mathbf{v} \times \mathbf{w}\| = \|\mathbf{v}\| \|\mathbf{w}\| \sin\theta$ and $\mathbf{v} \neq \mathbf{0}$, then:

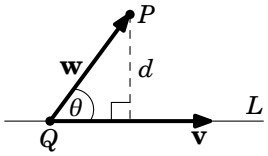


Figure 1.5.4

$$d = \frac{\|\mathbf{v} \times \mathbf{w}\|}{\|\mathbf{v}\|} \tag{1.23}$$

Example 1.21. Find the distance d from the point $P = (1,1,1)$ to the line L in Example 1.20.

Solution: From Example 1.20, we see that we can represent L in vector form as: $\mathbf{r} + t\mathbf{v}$, for $\mathbf{r} = (-3,1,-4)$ and $\mathbf{v} = (7,3,-2)$. Since the point $Q = (-3,1,-4)$ is on L, then for $\mathbf{w} = \overrightarrow{QP} = (1,1,1) - (-3,1,-4) = (4,0,5)$, we have:

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 7 & 3 & -2 \\ 4 & 0 & 5 \end{vmatrix} = \begin{vmatrix} 3 & -2 \\ 0 & 5 \end{vmatrix}\mathbf{i} - \begin{vmatrix} 7 & -2 \\ 4 & 5 \end{vmatrix}\mathbf{j} + \begin{vmatrix} 7 & 3 \\ 4 & 0 \end{vmatrix}\mathbf{k} = 15\mathbf{i} - 43\mathbf{j} - 12\mathbf{k}\text{ , so}$$

$$d = \frac{\|\mathbf{v} \times \mathbf{w}\|}{\|\mathbf{v}\|} = \frac{\left\|15\mathbf{i} - 43\mathbf{j} - 12\mathbf{k}\right\|}{\left\|(7,3,-2)\right\|} = \frac{\sqrt{15^2 + (-43)^2 + (-12)^2}}{\sqrt{7^2 + 3^2 + (-2)^2}} = \frac{\sqrt{2218}}{\sqrt{62}} = 5.98$$

It is clear that two lines L_1 and L_2, represented in vector form as $\mathbf{r}_1 + s\mathbf{v}_1$ and $\mathbf{r}_2 + t\mathbf{v}_2$, respectively, are parallel (denoted as $L_1 \parallel L_2$) if $\mathbf{v}_1$ and $\mathbf{v}_2$ are parallel. Also, L_1 and L_2 are perpendicular (denoted as $L_1 \perp L_2$) if $\mathbf{v}_1$ and $\mathbf{v}_2$ are perpendicular.

In 2-dimensional space, two lines are either identical, parallel, or they intersect. In 3-dimensional space, there is an additional possibility: two lines can be **skew**, that is, they do not intersect but they are not parallel. However, even though they are not parallel, skew lines are on parallel planes (see Figure 1.5.5).

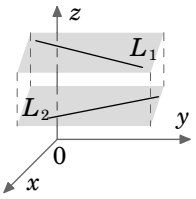


Figure 1.5.5

To determine whether two lines in $\mathbb{R}^3$ intersect, it is often easier to use the parametric representation of the lines. In this case, you should use different parameter variables (usually s and t) for the lines, since the values of the parameters may not be the same at the point of intersection. Setting the two (x,y,z) triples equal will result in a system of 3 equations in 2 unknowns (s and t).

Example 1.22. Find the point of intersection (if any) of the following lines:

$$\frac{x+1}{3} = \frac{y-2}{2} = \frac{z-1}{-1} \quad \text{and} \quad x+3 = \frac{y-8}{-3} = \frac{z+3}{2}$$

Solution: First we write the lines in parametric form, with parameters s and t:

$$x = -1 + 3s, \quad y = 2 + 2s, \quad z = 1 - s \quad \text{and} \quad x = -3 + t, \quad y = 8 - 3t, \quad z = -3 + 2t$$

The lines intersect when $(-1+3s, 2+2s, 1-s) = (-3+t, 8-3t, -3+2t)$ for some s, t:

$$\begin{aligned} -1 + 3s = -3 + t &: \Rightarrow t = 2 + 3s \\ 2 + 2s = 8 - 3t &: \Rightarrow 2 + 2s = 8 - 3(2+3s) = 2 - 9s \Rightarrow 2s = -9s \Rightarrow s = 0 \Rightarrow t = 2 + 3(0) = 2 \\ 1 - s = -3 + 2t &: \ 1 - 0 = -3 + 2(2) \Rightarrow 1 = 1 \ \checkmark \text{ (Note that we had to check this.)} \end{aligned}$$

Letting $s = 0$ in the equations for the first line, or letting $t = 2$ in the equations for the second line, gives the point of intersection $(-1,2,1)$.

We will now consider planes in 3-dimensional Euclidean space.

Plane through a point, perpendicular to a vector

Let P be a plane in $\mathbb{R}^3$, and suppose it contains a point $P_0 = (x_0, y_0, z_0)$. Let $\mathbf{n} = (a, b, c)$ be a nonzero vector which is perpendicular to the plane P. Such a vector is called a **normal vector** (or just a *normal*) to the plane. Now let (x, y, z) be any point in the plane P. Then the vector $\mathbf{r} = (x - x_0, y - y_0, z - z_0)$ lies in the plane P (see Figure 1.5.6). So if $\mathbf{r} \neq \mathbf{0}$, then $\mathbf{r} \perp \mathbf{n}$ and hence $\mathbf{n} \cdot \mathbf{r} = 0$. And if $\mathbf{r} = \mathbf{0}$ then we still have $\mathbf{n} \cdot \mathbf{r} = 0$.

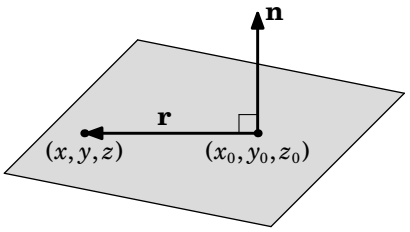


Figure 1.5.6 The plane P

Conversely, if (x, y, z) is any point in $\mathbb{R}^3$ such that $\mathbf{r} = (x - x_0, y - y_0, z - z_0) \neq \mathbf{0}$ and $\mathbf{n} \cdot \mathbf{r} = 0$, then $\mathbf{r} \perp \mathbf{n}$ and so (x, y, z) lies in P. This proves the following theorem:

Theorem 1.18. Let P be a plane in $\mathbb{R}^3$, let (x_0, y_0, z_0) be a point in P, and let $\mathbf{n} = (a, b, c)$ be a nonzero vector which is perpendicular to P. Then P consists of the points (x, y, z) satisfying the vector equation:

$$\mathbf{n} \cdot \mathbf{r} = 0 \tag{1.24}$$

where $\mathbf{r} = (x - x_0, y - y_0, z - z_0)$, or equivalently:

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \tag{1.25}$$

The above equation is called the **point-normal form** of the plane P.

Example 1.23. Find the equation of the plane P containing the point $(-3, 1, 3)$ and perpendicular to the vector $\mathbf{n} = (2, 4, 8)$.

Solution: By formula (1.25), the plane P consists of all points (x, y, z) such that:

$$2(x + 3) + 4(y - 1) + 8(z - 3) = 0$$

If we multiply out the terms in formula (1.25) and combine the constant terms, we get an equation of the plane in **normal form**:

$$ax + by + cz + d = 0 \tag{1.26}$$

For example, the normal form of the plane in Example 1.23 is $2x + 4y + 8z - 22 = 0$.

Plane containing three noncollinear points

In 2-dimensional and 3-dimensional space, two points determine a line. Two points do not determine a plane in $\mathbb{R}^3$. In fact, three *collinear* points (i.e. all on the same line) do not determine a plane; an infinite number of planes would contain the line on which those three points lie. However, three *noncollinear* points do determine a plane. For if Q, R and S are noncollinear points in $\mathbb{R}^3$, then $\overrightarrow{QR}$ and $\overrightarrow{QS}$ are nonzero vectors which are not parallel (by noncollinearity), and so their cross product $\overrightarrow{QR} \times \overrightarrow{QS}$ is perpendicular to both $\overrightarrow{QR}$ and $\overrightarrow{QS}$. So $\overrightarrow{QR}$ and $\overrightarrow{QS}$ (and hence Q, R and S) lie in the plane through the point Q with normal vector $\mathbf{n} = \overrightarrow{QR} \times \overrightarrow{QS}$ (see Figure 1.5.7).

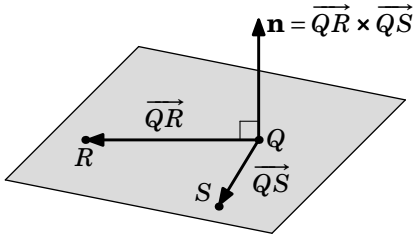


Figure 1.5.7 Noncollinear points Q, R, S

Example 1.24. Find the equation of the plane P containing the points $(2,1,3)$, $(1,-1,2)$ and $(3,2,1)$.

Solution: Let $Q = (2,1,3)$, $R = (1,-1,2)$ and $S = (3,2,1)$. Then for the vectors $\overrightarrow{QR} = (-1,-2,-1)$ and $\overrightarrow{QS} = (1,1,-2)$, the plane P has a normal vector

$$\mathbf{n} = \overrightarrow{QR} \times \overrightarrow{QS} = (-1,-2,-1) \times (1,1,-2) = (5,-3,1)$$

So using formula (1.25) with the point Q (we could also use R or S), the plane P consists of all points (x,y,z) such that:

$$5(x-2) - 3(y-1) + (z-3) = 0$$

or in normal form,

$$5x - 3y + z - 10 = 0$$

We mentioned earlier that skew lines in $\mathbb{R}^3$ lie on separate, parallel planes. So two skew lines do not determine a plane. But two (nonidentical) lines which either intersect or are parallel do determine a plane. In both cases, to find the equation of the plane that contains those two lines, simply pick from the two lines a total of three noncollinear points (i.e. one point from one line and two points from the other), then use the technique above, as in Example 1.24, to write the equation. We will leave examples of this as exercises for the reader.

Distance between a point and a plane

The distance between a point in $\mathbb{R}^3$ and a plane is the length of the line segment from that point to the plane which is perpendicular to the plane. The following theorem gives a formula for that distance.

Theorem 1.19. Let $Q = (x_0, y_0, z_0)$ be a point in $\mathbb{R}^3$, and let P be a plane with normal form $ax + by + cz + d = 0$ that does not contain Q. Then the distance D from Q to P is:

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}} \tag{1.27}$$

Proof: Let $R = (x, y, z)$ be any point in the plane P (so that $ax + by + cz + d = 0$) and let $\mathbf{r} = \overrightarrow{RQ} = (x_0 - x, y_0 - y, z_0 - z)$. Then $\mathbf{r} \neq \mathbf{0}$ since Q does not lie in P. From the normal form equation for P, we know that $\mathbf{n} = (a, b, c)$ is a normal vector for P. Now, any plane divides $\mathbb{R}^3$ into two disjoint parts. Assume that $\mathbf{n}$ points toward the side of P where the point Q is located. Place $\mathbf{n}$ so that its initial point is at R, and let θ be the angle between $\mathbf{r}$ and $\mathbf{n}$. Then $0° < \theta < 90°$, so $\cos\theta > 0$. Thus, the distance D is $\cos\theta \, \|\mathbf{r}\| = |\cos\theta| \, \|\mathbf{r}\|$ (see Figure 1.5.8).

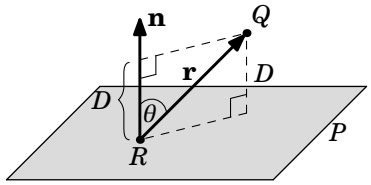


Figure 1.5.8

By Theorem 1.6 in Section 1.3, we know that $\cos\theta = \dfrac{\mathbf{n} \cdot \mathbf{r}}{\|\mathbf{n}\| \, \|\mathbf{r}\|}$, so

$$D = |\cos\theta| \, \|\mathbf{r}\| = \frac{|\mathbf{n} \cdot \mathbf{r}|}{\|\mathbf{n}\| \, \|\mathbf{r}\|} \|\mathbf{r}\| = \frac{|\mathbf{n} \cdot \mathbf{r}|}{\|\mathbf{n}\|} = \frac{|a(x_0 - x) + b(y_0 - y) + c(z_0 - z)|}{\sqrt{a^2 + b^2 + c^2}}$$
$$= \frac{|ax_0 + by_0 + cz_0 - (ax + by + cz)|}{\sqrt{a^2 + b^2 + c^2}} = \frac{|ax_0 + by_0 + cz_0 - (-d)|}{\sqrt{a^2 + b^2 + c^2}} = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}$$

If $\mathbf{n}$ points away from the side of P where the point Q is located, then $90° < \theta < 180°$ and so $\cos\theta < 0$. The distance D is then $|\cos\theta| \, \|\mathbf{r}\|$, and thus repeating the same argument as above still gives the same result. **QED**

Example 1.25. Find the distance D from $(2, 4, -5)$ to the plane from Example 1.24.

Solution: Recall that the plane is given by $5x - 3y + z - 10 = 0$. So

$$D = \frac{|5(2) - 3(4) + 1(-5) - 10|}{\sqrt{5^2 + (-3)^2 + 1^2}} = \frac{|-17|}{\sqrt{35}} = \frac{17}{\sqrt{35}} \approx 2.87$$

Line of intersection of two planes

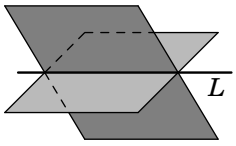


Figure 1.5.9

Note that two planes are parallel if they have normal vectors that are parallel, and the planes are perpendicular if their normal vectors are perpendicular. If two planes do intersect, they do so in a line (see Figure 1.5.9). Suppose that two planes P_1 and P_2 with normal vectors $\mathbf{n}_1$ and $\mathbf{n}_2$, respectively, intersect in a line L. Since $\mathbf{n}_1 \times \mathbf{n}_2 \perp \mathbf{n}_1$, then $\mathbf{n}_1 \times \mathbf{n}_2$ is parallel to the plane P_1. Likewise, $\mathbf{n}_1 \times \mathbf{n}_2 \perp \mathbf{n}_2$ means that $\mathbf{n}_1 \times \mathbf{n}_2$ is also parallel to P_2. Thus, $\mathbf{n}_1 \times \mathbf{n}_2$ is parallel to the intersection of P_1 and P_2, i.e. $\mathbf{n}_1 \times \mathbf{n}_2$ is parallel to L. Thus, we can write L in the following vector form:

$$L : \mathbf{r} + t(\mathbf{n}_1 \times \mathbf{n}_2), \text{ for } -\infty < t < \infty \tag{1.28}$$

where $\mathbf{r}$ is any vector pointing to a point belonging to both planes. To find a point in both planes, find a common solution (x, y, z) to the two normal form equations of the planes. This can often be made easier by setting one of the coordinate variables to zero, which leaves you to solve two equations in just two unknowns.

Example 1.26. Find the line of intersection L of the planes $5x - 3y + z - 10 = 0$ and $2x + 4y - z + 3 = 0$.

Solution: The plane $5x - 3y + z - 10 = 0$ has normal vector $\mathbf{n}_1 = (5, -3, 1)$ and the plane $2x + 4y - z + 3 = 0$ has normal vector $\mathbf{n}_2 = (2, 4, -1)$. Since $\mathbf{n}_1$ and $\mathbf{n}_2$ are not scalar multiples, then the two planes are not parallel and hence will intersect. A point (x, y, z) on both planes will satisfy the following system of two equations in three unknowns:

$$\begin{aligned} 5x - 3y + z - 10 &= 0 \\ 2x + 4y - z + 3 &= 0 \end{aligned}$$

Set $x = 0$ (why is that a good choice?). Then the above equations are reduced to:

$$\begin{aligned} -3y + z - 10 &= 0 \\ 4y - z + 3 &= 0 \end{aligned}$$

The second equation gives $z = 4y + 3$, substituting that into the first equation gives $y = 7$. Then $z = 31$, and so the point $(0, 7, 31)$ is on L. Since $\mathbf{n}_1 \times \mathbf{n}_2 = (-1, 7, 26)$, then L is given by:

$$\mathbf{r} + t(\mathbf{n}_1 \times \mathbf{n}_2) = (0, 7, 31) + t(-1, 7, 26), \text{ for } -\infty < t < \infty$$

or in parametric form:

$$x = -t, \quad y = 7 + 7t, \quad z = 31 + 26t, \text{ for } -\infty < t < \infty$$

Exercises

A

For Exercises 1-4, write the line L through the point P and parallel to the vector $\mathbf{v}$ in the following forms: (a) vector, (b) parametric, and (c) symmetric.

1. $P = (2,3,-2)$, $\mathbf{v} = (5,4,-3)$

2. $P = (3,-1,2)$, $\mathbf{v} = (2,8,1)$

3. $P = (2,1,3)$, $\mathbf{v} = (1,0,1)$

4. $P = (0,0,0)$, $\mathbf{v} = (7,2,-10)$

For Exercises 5-6, write the line L through the points P_1 and P_2 in parametric form.

5. $P_1 = (1,-2,-3)$, $P_2 = (3,5,5)$

6. $P_1 = (4,1,5)$, $P_2 = (-2,1,3)$

For Exercises 7-8, find the distance d from the point P to the line L.

7. $P = (1,-1,-1)$, $L : x = -2-2t$, $y = 4t$, $z = 7+t$

8. $P = (0,0,0)$, $L : x = 3+2t$, $y = 4+3t$, $z = 5+4t$

For Exercises 9-10, find the point of intersection (if any) of the given lines.

9. $x = 7+3s$, $y = -4-3s$, $z = -7-5s$ and $x = 1+6t$, $y = 2+t$, $z = 3-2t$

10. $\dfrac{x-6}{4} = y+3 = z$ and $\dfrac{x-11}{3} = \dfrac{y-14}{-6} = \dfrac{z+9}{2}$

For Exercises 11-12, write the normal form of the plane P containing the point Q and perpendicular to the vector $\mathbf{n}$.

11. $Q = (5,1,-2)$, $\mathbf{n} = (4,-4,3)$

12. $Q = (6,-2,0)$, $\mathbf{n} = (2,6,4)$

For Exercises 13-14, write the normal form of the plane containing the given points.

13. $(1,0,3)$, $(1,2,-1)$, $(6,1,6)$

14. $(-3,1,-3)$, $(4,-4,3)$, $(0,0,1)$

15. Write the normal form of the plane containing the lines from Exercise 9.

16. Write the normal form of the plane containing the lines from Exercise 10.

For Exercises 17-18, find the distance D from the point Q to the plane P.

17. $Q = (4,1,2)$, $P : 3x-y-5z+8 = 0$

18. $Q = (0,2,0)$, $P : -5x+2y-7z+1 = 0$

For Exercises 19-20, find the line of intersection (if any) of the given planes.

19. $x+3y+2z-6 = 0$, $2x-y+z+2 = 0$

20. $3x+y-5z = 0$, $x+2y+z+4 = 0$

B

21. Find the point(s) of intersection (if any) of the line $\dfrac{x-6}{4} = y+3 = z$ with the plane $x+3y+2z-6 = 0$. (*Hint: Put the equations of the line into the equation of the plane.*)

1.6 Surfaces

In the previous section we discussed planes in Euclidean space. A plane is an example of a *surface*, which we will define informally[8] as the solution set of the equation $F(x,y,z) = 0$ in $\mathbb{R}^3$, for some real-valued function F. For example, a plane given by $ax + by + cz + d = 0$ is the solution set of $F(x,y,z) = 0$ for the function $F(x,y,z) = ax + by + cz + d$. Surfaces are 2-dimensional. The plane is the simplest surface, since it is "flat". In this section we will look at some surfaces that are more complex, the most important of which are the sphere and the cylinder.

Definition 1.9. A **sphere** S is the set of all points (x,y,z) in $\mathbb{R}^3$ which are a fixed distance r (called the **radius**) from a fixed point $P_0 = (x_0, y_0, z_0)$ (called the **center** of the sphere):

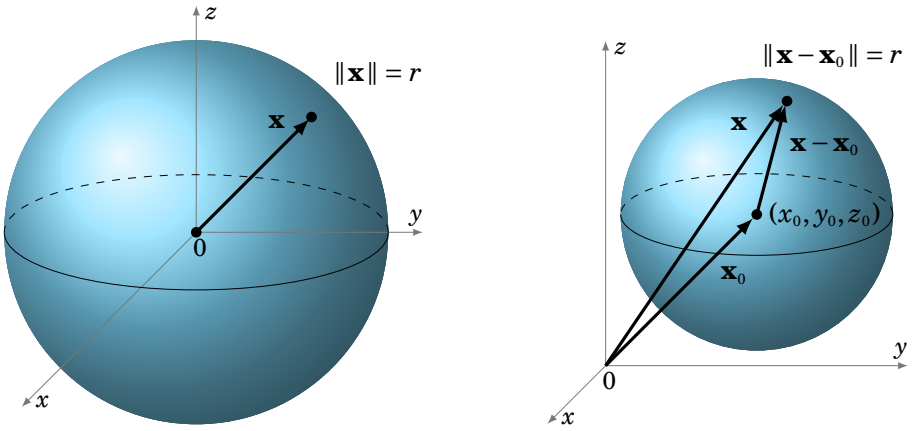
$$S = \{(x,y,z) : (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 = r^2\} \tag{1.29}$$

Using vector notation, this can be written in the equivalent form:

$$S = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| = r\} \tag{1.30}$$

where $\mathbf{x} = (x,y,z)$ and $\mathbf{x}_0 = (x_0, y_0, z_0)$ are vectors.

Figure 1.6.1 illustrates the vectorial approach to spheres.



(a) radius r, center $(0,0,0)$ **(b)** radius r, center (x_0, y_0, z_0)

Figure 1.6.1 Spheres in $\mathbb{R}^3$

Note in Figure 1.6.1(a) that the intersection of the sphere with the xy-plane is a circle of radius r (i.e. a *great circle*, given by $x^2 + y^2 = r^2$ as a subset of $\mathbb{R}^2$). Similarly for the intersections with the xz-plane and the yz-plane. In general, a plane intersects a sphere either at a single point or in a circle.

[8] See O'NEILL for a deeper and more rigorous discussion of surfaces.

Example 1.27. Find the intersection of the sphere $x^2+y^2+z^2=169$ with the plane $z=12$.

Solution: The sphere is centered at the origin and has radius $13=\sqrt{169}$, so it does intersect the plane $z=12$. Putting $z=12$ into the equation of the sphere gives

$$x^2+y^2+12^2=169$$
$$x^2+y^2=169-144=25=5^2$$

which is a circle of radius 5 centered at $(0,0,12)$, parallel to the xy-plane (see Figure 1.6.2).

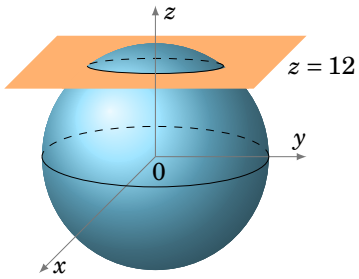


Figure 1.6.2

If the equation in formula (1.29) is multiplied out, we get an equation of the form:

$$x^2+y^2+z^2+ax+by+cz+d=0 \tag{1.31}$$

for some constants a, b, c and d. Conversely, an equation of this form *may* describe a sphere, which can be determined by completing the square for the x, y and z variables.

Example 1.28. Is $2x^2+2y^2+2z^2-8x+4y-16z+10=0$ the equation of a sphere?

Solution: Dividing both sides of the equation by 2 gives

$$x^2+y^2+z^2-4x+2y-8z+5=0$$
$$(x^2-4x+4)+(y^2+2y+1)+(z^2-8z+16)+5-4-1-16=0$$
$$(x-2)^2+(y+1)^2+(z-4)^2=16$$

which is a sphere of radius 4 centered at $(2,-1,4)$.

Example 1.29. Find the points(s) of intersection (if any) of the sphere from Example 1.28 and the line $x=3+t$, $y=1+2t$, $z=3-t$.

Solution: Put the equations of the line into the equation of the sphere, which was $(x-2)^2+(y+1)^2+(z-4)^2=16$, and solve for t:

$$(3+t-2)^2+(1+2t+1)^2+(3-t-4)^2=16$$
$$(t+1)^2+(2t+2)^2+(-t-1)^2=16$$
$$6t^2+12t-10=0$$

The quadratic formula gives the solutions $t=-1\pm\dfrac{4}{\sqrt{6}}$. Putting those two values into the equations of the line gives the following two points of intersection:

$$\left(2+\frac{4}{\sqrt{6}},-1+\frac{8}{\sqrt{6}},4-\frac{4}{\sqrt{6}}\right) \quad \text{and} \quad \left(2-\frac{4}{\sqrt{6}},-1-\frac{8}{\sqrt{6}},4+\frac{4}{\sqrt{6}}\right)$$

If two spheres intersect, they do so either at a single point or in a circle.

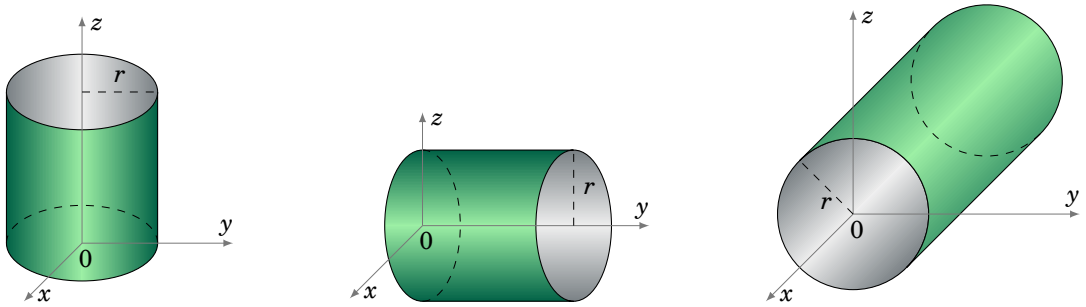
Example 1.30. Find the intersection (if any) of the spheres $x^2+y^2+z^2=25$ and $x^2+y^2+(z-2)^2=16$.

Solution: For any point (x,y,z) on both spheres, we see that

$$\begin{aligned} x^2+y^2+z^2=25 &\Rightarrow x^2+y^2=25-z^2, \text{ and} \\ x^2+y^2+(z-2)^2=16 &\Rightarrow x^2+y^2=16-(z-2)^2, \text{ so} \\ 16-(z-2)^2=25-z^2 &\Rightarrow 4z-4=9 \Rightarrow z=13/4 \\ &\Rightarrow x^2+y^2=25-(13/4)^2=231/16 \end{aligned}$$

$\therefore$ The intersection is the circle $x^2+y^2=\frac{231}{16}$ of radius $\frac{\sqrt{231}}{4}\approx 3.8$ centered at $(0,0,\frac{13}{4})$.

The cylinders that we will consider are *right circular cylinders*. These are cylinders obtained by moving a line L along a circle C in $\mathbb{R}^3$ in a way so that L is always perpendicular to the plane containing C. We will only consider the cases where the plane containing C is parallel to one of the three coordinate planes (see Figure 1.6.3).



(a) $x^2+y^2=r^2$, any z **(b)** $x^2+z^2=r^2$, any y **(c)** $y^2+z^2=r^2$, any x

Figure 1.6.3 Cylinders in $\mathbb{R}^3$

For example, the equation of a cylinder whose base circle C lies in the xy-plane and is centered at $(a,b,0)$ and has radius r is

$$(x-a)^2+(y-b)^2=r^2, \tag{1.32}$$

where the value of the z coordinate is unrestricted. Similar equations can be written when the base circle lies in one of the other coordinate planes. A plane intersects a right circular cylinder in a circle, ellipse, or one or two lines, depending on whether that plane is parallel, oblique[9], or perpendicular, respectively, to the plane containing C. The intersection of a surface with a plane is called the **trace** of the surface.

[9] i.e. at an angle strictly between 0° and 90°.

The equations of spheres and cylinders are examples of *second-degree equations* in $\mathbb{R}^3$, i.e. equations of the form

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0 \tag{1.33}$$

for some constants $A, B, \ldots, J$. If the above equation is not that of a sphere, cylinder, plane, line or point, then the resulting surface is called a **quadric surface**.

One type of quadric surface is the **ellipsoid**, given by an equation of the form:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{1.34}$$

In the case where $a = b = c$, this is just a sphere. In general, an ellipsoid is egg-shaped (think of an ellipse rotated around its major axis). Its traces in the coordinate planes are ellipses.

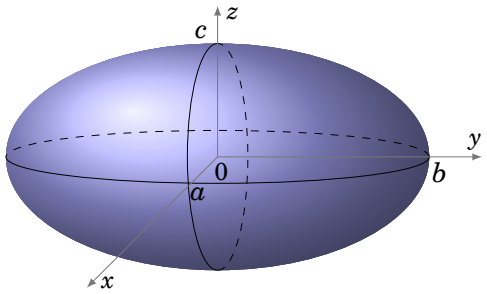


Figure 1.6.4 Ellipsoid

Two other types of quadric surfaces are the **hyperboloid of one sheet**, given by an equation of the form:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \tag{1.35}$$

and the **hyperboloid of two sheets**, whose equation has the form:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \tag{1.36}$$

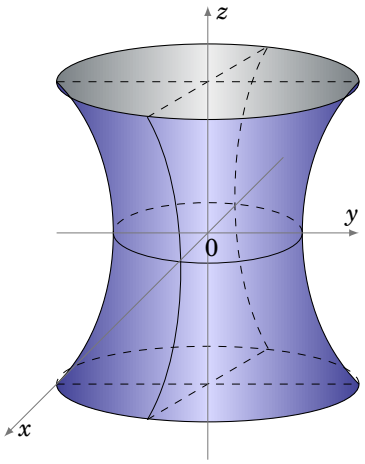


Figure 1.6.5 Hyperboloid of one sheet

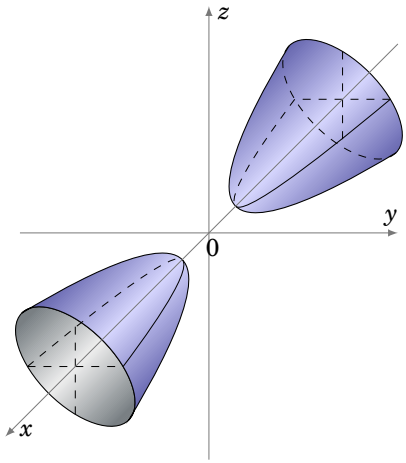


Figure 1.6.6 Hyperboloid of two sheets

For the hyperboloid of one sheet, the trace in any plane parallel to the xy-plane is an ellipse. The traces in the planes parallel to the xz- or yz-planes are hyperbolas (see Figure 1.6.5), except for the special cases $x = \pm a$ and $y = \pm b$; in those planes the traces are pairs of intersecting lines (see Exercise 8).

For the hyperboloid of two sheets, the trace in any plane parallel to the xy- or xz-plane is a hyperbola (see Figure 1.6.6). There is no trace in the yz-plane. In any plane parallel to the yz-plane for which $|x| > |a|$, the trace is an ellipse.

The **elliptic paraboloid** is another type of quadric surface, whose equation has the form:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z}{c} \tag{1.37}$$

The traces in planes parallel to the xy-plane are ellipses, though in the xy-plane itself the trace is a single point. The traces in planes parallel to the xz- or yz-planes are parabolas. Figure 1.6.7 shows the case where $c > 0$. When $c < 0$ the surface is turned downward. In the case where $a = b$, the surface is called a *paraboloid of revolution*, which is often used as a reflecting surface, e.g. in vehicle headlights.[10]

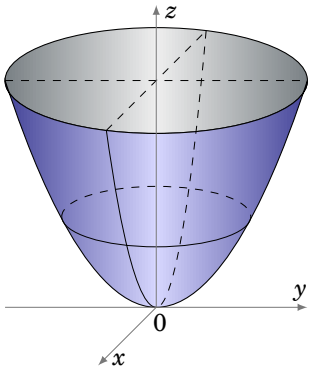


Figure 1.6.7 Paraboloid

A more complicated quadric surface is the **hyperbolic paraboloid**, given by:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \frac{z}{c} \tag{1.38}$$

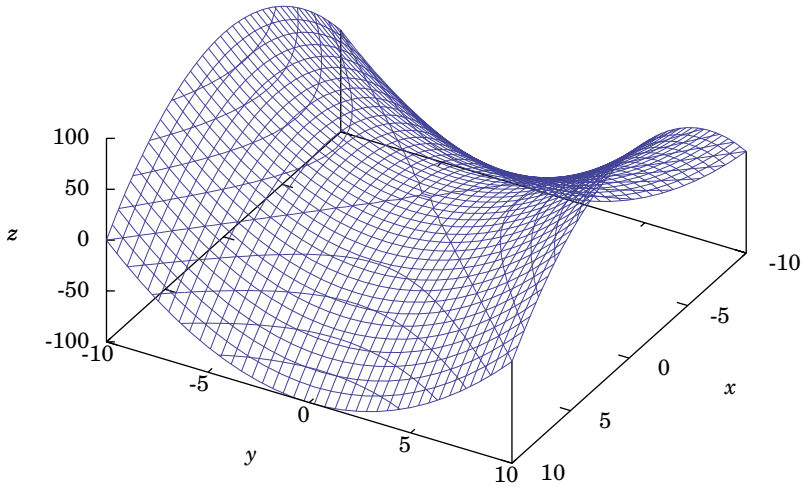


Figure 1.6.8 Hyperbolic paraboloid

[10] For a discussion of this see pp. 157-158 in HECHT.

The hyperbolic paraboloid can be tricky to draw; using graphing software on a computer can make it easier. For example, Figure 1.6.8 was created using the free Gnuplot package (see Appendix C). It shows the graph of the hyperbolic paraboloid $z = y^2 - x^2$, which is the special case where $a = b = 1$ and $c = -1$ in equation (1.38). The mesh lines on the surface are the traces in planes parallel to the coordinate planes. So we see that the traces in planes parallel to the xz-plane are parabolas pointing upward, while the traces in planes parallel to the yz-plane are parabolas pointing downward. Also, notice that the traces in planes parallel to the xy-plane are hyperbolas, though in the xy-plane itself the trace is a pair of intersecting lines through the origin. This is true in general when $c < 0$ in equation (1.38). When $c > 0$, the surface would be similar to that in Figure 1.6.8, only rotated 90° around the z-axis and the nature of the traces in planes parallel to the xz- or yz-planes would be reversed.

The last type of quadric surface that we will consider is the **elliptic cone**, which has an equation of the form:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0 \qquad (1.39)$$

The traces in planes parallel to the xy-plane are ellipses, except in the xy-plane itself where the trace is a single point. The traces in planes parallel to the xz- or yz-planes are hyperbolas, except in the xz- and yz-planes themselves where the traces are pairs of intersecting lines.

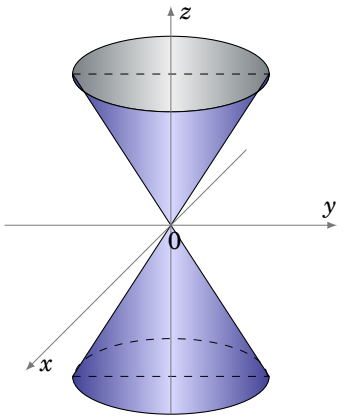


Figure 1.6.9 Elliptic cone

Notice that every point on the elliptic cone is on a line which lies entirely on the surface; in Figure 1.6.9 these lines all go through the origin. This makes the elliptic cone an example of a *ruled surface*. The cylinder is also a ruled surface.

What may not be as obvious is that both the hyperboloid of one sheet and the hyperbolic paraboloid are ruled surfaces. In fact, on both surfaces there are *two* lines through each point on the surface (see Exercises 11-12). Such surfaces are called *doubly ruled surfaces*, and the pairs of lines are called a *regulus*.

It is clear that for each of the six types of quadric surfaces that we discussed, the surface can be translated away from the origin (e.g. by replacing x^2 by $(x-x_0)^2$ in its equation). It can be proved[11] that *every* quadric surface can be translated and/or rotated so that its equation matches one of the six types that we described. For example, $z = 2xy$ is a case of equation (1.33) with "mixed" variables, e.g. with $D \neq 0$ so that we get an xy term. This equation does not match any of the types we considered. However, by rotating the x- and y-axes by 45° in the xy-plane by means of the coordinate transformation $x = (x' - y')/\sqrt{2}$, $y = (x' + y')/\sqrt{2}$, $z = z'$, then $z = 2xy$ becomes the hyperbolic paraboloid $z' = (x')^2 - (y')^2$ in the (x', y', z') coordinate system. That is, $z = 2xy$ is a hyperbolic paraboloid as in equation (1.38), but rotated 45° in the xy-plane.

[11] See Ch. 7 in POGORELOV.

Exercises

A

For Exercises 1-4, determine if the given equation describes a sphere. If so, find its radius and center.

1. $x^2+y^2+z^2-4x-6y-10z+37=0$

2. $x^2+y^2+z^2+2x-2y-8z+19=0$

3. $2x^2+2y^2+2z^2+4x+4y+4z-44=0$

4. $x^2+y^2-z^2+12x+2y-4z+32=0$

5. Find the point(s) of intersection of the sphere $(x-3)^2+(y+1)^2+(z-3)^2=9$ and the line $x=-1+2t$, $y=-2-3t$, $z=3+t$.

B

6. Find the intersection of the spheres $x^2+y^2+z^2=9$ and $(x-4)^2+(y+2)^2+(z-4)^2=9$.

7. Find the intersection of the sphere $x^2+y^2+z^2=9$ and the cylinder $x^2+y^2=4$.

8. Find the trace of the hyperboloid of one sheet $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$ in the plane $x=a$, and the trace in the plane $y=b$.

9. Find the trace of the hyperbolic paraboloid $\frac{x^2}{a^2}-\frac{y^2}{b^2}=\frac{z}{c}$ in the xy-plane.

C

10. It can be shown that any four *noncoplanar* points (i.e. points that do not lie in the same plane) determine a sphere.[12] Find the equation of the sphere that passes through the points $(0,0,0)$, $(0,0,2)$, $(1,-4,3)$ and $(0,-1,3)$. *(Hint: Equation (1.31))*

11. Show that the hyperboloid of one sheet is a doubly ruled surface, i.e. each point on the surface is on two lines lying entirely on the surface. *(Hint: Write equation (1.35) as* $\frac{x^2}{a^2}-\frac{z^2}{c^2}=1-\frac{y^2}{b^2}$ *, factor each side. Recall that two planes intersect in a line.)*

12. Show that the hyperbolic paraboloid is a doubly ruled surface. *(Hint: Exercise 11)*

13. Let S be the sphere with radius 1 centered at $(0,0,1)$, and let S^* be S without the "north pole" point $(0,0,2)$. Let (a,b,c) be an arbitrary point on S^*. Then the line passing through $(0,0,2)$ and (a,b,c) intersects the xy-plane at some point $(x,y,0)$, as in Figure 1.6.10. Find this point $(x,y,0)$ in terms of a, b and c.
(Note: Every point in the xy-plane can be matched with a point on S^*, and vice versa, in this manner. This method is called *stereographic projection*, which essentially identifies all of $\mathbb{R}^2$ with a "punctured" sphere.)

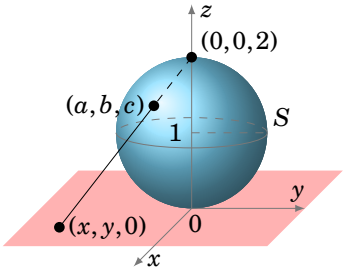


Figure 1.6.10

[12] See WELCHONS and KRICKENBERGER, p. 160, for a proof.

1.7 Curvilinear Coordinates

The Cartesian coordinates of a point (x,y,z) are determined by following straight paths starting from the origin: first along the x-axis, then parallel to the y-axis, then parallel to the z-axis, as in Figure 1.7.1. In *curvilinear coordinate systems*, these paths can be curved. The two types of curvilinear coordinates which we will consider are cylindrical and spherical coordinates. Instead of referencing a point in terms of sides of a rectangular parallelepiped, as with Cartesian coordinates, we will think of the point as lying on a cylinder or sphere. Cylindrical coordinates are often used when there is symmetry around the z-axis; spherical coordinates are useful when there is symmetry about the origin.

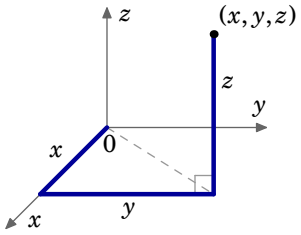


Figure 1.7.1

Let $P = (x,y,z)$ be a point in Cartesian coordinates in $\mathbb{R}^3$, and let $P_0 = (x,y,0)$ be the projection of P upon the xy-plane. Treating (x,y) as a point in $\mathbb{R}^2$, let (r,θ) be its polar coordinates (see Figure 1.7.2). Let ρ be the length of the line segment from the origin to P, and let ϕ be the angle between that line segment and the positive z-axis (see Figure 1.7.3). ϕ is called the *zenith angle*. Then the **cylindrical coordinates** (r,θ,z) and the **spherical coordinates** (ρ,θ,ϕ) of $P(x,y,z)$ are defined as follows:[13]

Cylindrical coordinates (r,θ,z):

$$x = r\cos\theta \qquad r = \sqrt{x^2+y^2}$$
$$y = r\sin\theta \qquad \theta = \tan^{-1}\left(\tfrac{y}{x}\right)$$
$$z = z \qquad z = z$$

where $0 \le \theta \le \pi$ if $y \ge 0$ and $\pi < \theta < 2\pi$ if $y < 0$

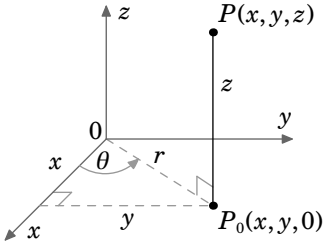


Figure 1.7.2
Cylindrical coordinates

Spherical coordinates (ρ,θ,ϕ):

$$x = \rho\sin\phi\cos\theta \qquad \rho = \sqrt{x^2+y^2+z^2}$$
$$y = \rho\sin\phi\sin\theta \qquad \theta = \tan^{-1}\left(\tfrac{y}{x}\right)$$
$$z = \rho\cos\phi \qquad \phi = \cos^{-1}\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right)$$

where $0 \le \theta \le \pi$ if $y \ge 0$ and $\pi < \theta < 2\pi$ if $y < 0$

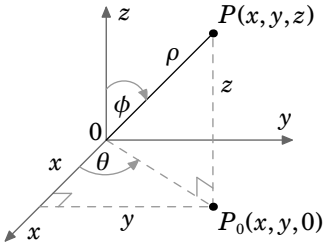


Figure 1.7.3
Spherical coordinates

Both θ and ϕ are measured in radians. Note that $r \ge 0$, $0 \le \theta < 2\pi$, $\rho \ge 0$ and $0 \le \phi \le \pi$. Also, θ is undefined when $(x,y) = (0,0)$, and ϕ is undefined when $(x,y,z) = (0,0,0)$.

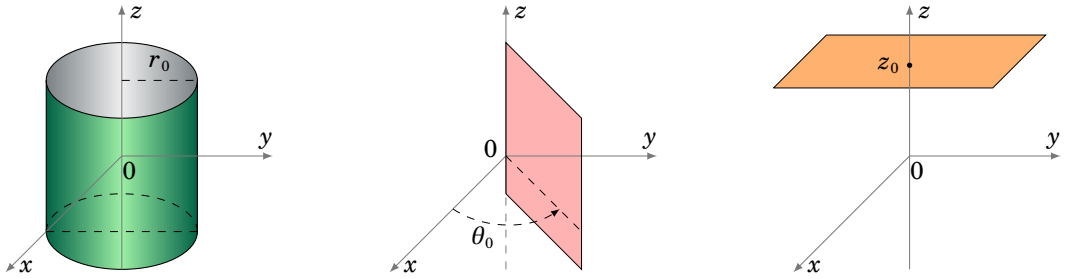
[13] This "standard" definition of spherical coordinates used by mathematicians results in a left-handed system. For this reason, physicists usually switch the definitions of θ and ϕ to make (ρ,θ,ϕ) a right-handed system.

Example 1.31. Convert the point $(-2,-2,1)$ from Cartesian coordinates to (a) cylindrical and (b) spherical coordinates.

Solution: (a) $r = \sqrt{(-2)^2+(-2)^2} = 2\sqrt{2}$, $\theta = \tan^{-1}\left(\frac{-2}{-2}\right) = \tan^{-1}(1) = \frac{5\pi}{4}$, since $y = -2 < 0$.
$\therefore\ (r,\theta,z) = \left(2\sqrt{2}, \frac{5\pi}{4}, 1\right)$

(b) $\rho = \sqrt{(-2)^2+(-2)^2+1^2} = \sqrt{9} = 3$, $\phi = \cos^{-1}\left(\frac{1}{3}\right) \approx 1.23$ radians.
$\therefore\ (\rho,\theta,\phi) = \left(3, \frac{5\pi}{4}, 1.23\right)$

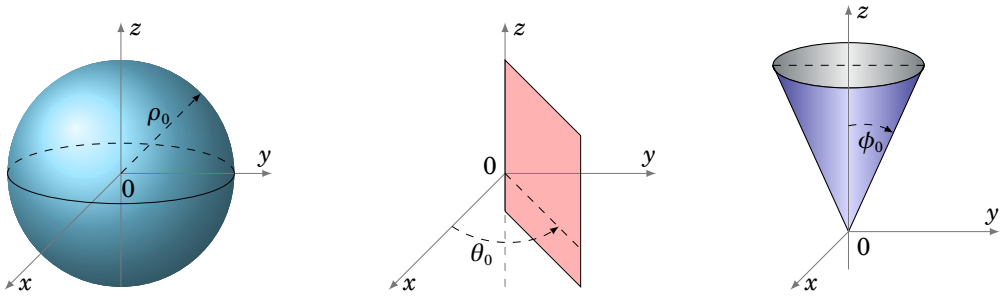
For cylindrical coordinates (r,θ,z), and constants r_0, θ_0 and z_0, we see from Figure 1.7.4 that the surface $r = r_0$ is a cylinder of radius r_0 centered along the z-axis, the surface $\theta = \theta_0$ is a half-plane emanating from the z-axis, and the surface $z = z_0$ is a plane parallel to the xy-plane.



(a) $r = r_0$ **(b)** $\theta = \theta_0$ **(c)** $z = z_0$

Figure 1.7.4 Cylindrical coordinate surfaces

For spherical coordinates (ρ,θ,ϕ), and constants ρ_0, θ_0 and ϕ_0, we see from Figure 1.7.5 that the surface $\rho = \rho_0$ is a sphere of radius ρ_0 centered at the origin, the surface $\theta = \theta_0$ is a half-plane emanating from the z-axis, and the surface $\phi = \phi_0$ is a circular cone whose vertex is at the origin.



(a) $\rho = \rho_0$ **(b)** $\theta = \theta_0$ **(c)** $\phi = \phi_0$

Figure 1.7.5 Spherical coordinate surfaces

Figures 1.7.4(a) and 1.7.5(a) show how these coordinate systems got their names.

Sometimes the equation of a surface in Cartesian coordinates can be transformed into a simpler equation in some other coordinate system, as in the following example.

Example 1.32. Write the equation of the cylinder $x^2 + y^2 = 4$ in cylindrical coordinates.

Solution: Since $r = \sqrt{x^2 + y^2}$, then the equation in cylindrical coordinates is $r = 2$.

Using spherical coordinates to write the equation of a sphere does not necessarily make the equation simpler, if the sphere is not centered at the origin.

Example 1.33. Write the equation $(x-2)^2 + (y-1)^2 + z^2 = 9$ in spherical coordinates.

Solution: Multiplying the equation out gives

$$x^2 + y^2 + z^2 - 4x - 2y + 5 = 9 \text{ , so we get}$$
$$\rho^2 - 4\rho\sin\phi\cos\theta - 2\rho\sin\phi\sin\theta - 4 = 0 \text{ , or}$$
$$\rho^2 - 2\sin\phi(2\cos\theta - \sin\theta)\rho - 4 = 0$$

after combining terms. Note that this actually makes it more difficult to figure out what the surface is, as opposed to the Cartesian equation where you could immediately identify the surface as a sphere of radius 3 centered at $(2, 1, 0)$.

Example 1.34. Describe the surface given by $\theta = z$ in cylindrical coordinates.

Solution: This surface is called a *helicoid*. As the (vertical) z coordinate increases, so does the angle θ, while the radius r is unrestricted. So this sweeps out a (ruled!) surface shaped like a spiral staircase, where the spiral has an infinite radius. Figure 1.7.6 shows a section of this surface restricted to $0 \le z \le 4\pi$ and $0 \le r \le 2$.

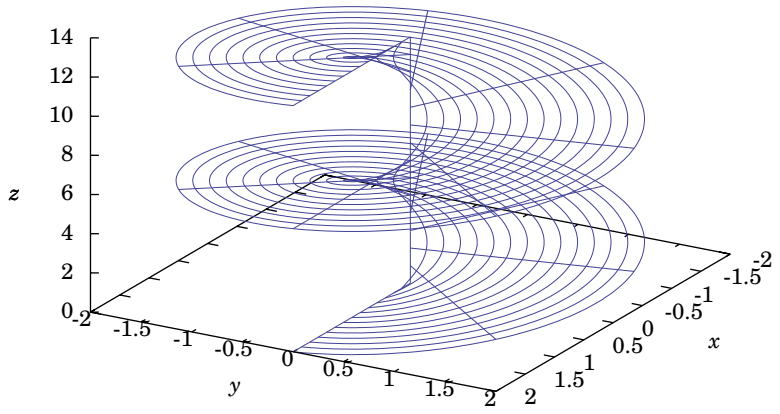


Figure 1.7.6 Helicoid $\theta = z$

Exercises

A

For Exercises 1-4, find the (a) cylindrical and (b) spherical coordinates of the point whose Cartesian coordinates are given.

1. $(2, 2\sqrt{3}, -1)$ **2.** $(-5, 5, 6)$ **3.** $(\sqrt{21}, -\sqrt{7}, 0)$ **4.** $(0, \sqrt{2}, 2)$

For Exercises 5-7, write the given equation in (a) cylindrical and (b) spherical coordinates.

5. $x^2 + y^2 + z^2 = 25$ **6.** $x^2 + y^2 = 2y$ **7.** $x^2 + y^2 + 9z^2 = 36$

B

8. Describe the intersection of the surfaces whose equations in spherical coordinates are $\theta = \frac{\pi}{2}$ and $\phi = \frac{\pi}{4}$.

9. Show that for $a \neq 0$, the equation $\rho = 2a \sin\phi \cos\theta$ in spherical coordinates describes a sphere centered at $(a, 0, 0)$ with radius $|a|$.

C

10. Let $P = (a, \theta, \phi)$ be a point in spherical coordinates, with $a > 0$ and $0 < \phi < \pi$. Then P lies on the sphere $\rho = a$. Since $0 < \phi < \pi$, the line segment from the origin to P can be extended to intersect the cylinder given by $r = a$ (in cylindrical coordinates). Find the cylindrical coordinates of that point of intersection.

11. Let P_1 and P_2 be points whose spherical coordinates are $(\rho_1, \theta_1, \phi_1)$ and $(\rho_2, \theta_2, \phi_2)$, respectively. Let $\mathbf{v}_1$ be the vector from the origin to P_1, and let $\mathbf{v}_2$ be the vector from the origin to P_2. For the angle γ between $\mathbf{v}_1$ and $\mathbf{v}_2$, show that

$$\cos\gamma = \cos\phi_1 \cos\phi_2 + \sin\phi_1 \sin\phi_2 \cos(\theta_2 - \theta_1).$$

This formula is used in electrodynamics to prove the addition theorem for spherical harmonics, which provides a general expression for the electrostatic potential at a point due to a unit charge. See pp. 100-102 in JACKSON.

12. Show that the distance d between the points P_1 and P_2 with cylindrical coordinates (r_1, θ_1, z_1) and (r_2, θ_2, z_2), respectively, is

$$d = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_2 - \theta_1) + (z_2 - z_1)^2}.$$

13. Show that the distance d between the points P_1 and P_2 with spherical coordinates $(\rho_1, \theta_1, \phi_1)$ and $(\rho_2, \theta_2, \phi_2)$, respectively, is

$$d = \sqrt{\rho_1^2 + \rho_2^2 - 2\rho_1 \rho_2 [\sin\phi_1 \sin\phi_2 \cos(\theta_2 - \theta_1) + \cos\phi_1 \cos\phi_2]}.$$

1.8 Vector-Valued Functions

Now that we are familiar with vectors and their operations, we can begin discussing functions whose values are vectors.

Definition 1.10. A **vector-valued function of a real variable** is a rule that associates a vector $\mathbf{f}(t)$ with a real number t, where t is in some subset D of $\mathbb{R}^1$ (called the **domain** of $\mathbf{f}$). We write $\mathbf{f}: D \to \mathbb{R}^3$ to denote that $\mathbf{f}$ is a mapping of D into $\mathbb{R}^3$.

For example, $\mathbf{f}(t) = t\mathbf{i} + t^2\mathbf{j} + t^3\mathbf{k}$ is a vector-valued function in $\mathbb{R}^3$, defined for all real numbers t. We would write $\mathbf{f}: \mathbb{R} \to \mathbb{R}^3$. At $t = 1$ the value of the function is the vector $\mathbf{i} + \mathbf{j} + \mathbf{k}$, which in Cartesian coordinates has the terminal point $(1, 1, 1)$.

A vector-valued function of a real variable can be written in component form as

$$\mathbf{f}(t) = f_1(t)\mathbf{i} + f_2(t)\mathbf{j} + f_3(t)\mathbf{k}$$

or in the form

$$\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$$

for some real-valued functions $f_1(t)$, $f_2(t)$, $f_3(t)$, called the *component functions* of $\mathbf{f}$. The first form is often used when emphasizing that $\mathbf{f}(t)$ is a vector, and the second form is useful when considering just the terminal points of the vectors. By identifying vectors with their terminal points, a curve in space can be written as a vector-valued function.

Example 1.35. Define $\mathbf{f}: \mathbb{R} \to \mathbb{R}^3$ by $\mathbf{f}(t) = (\cos t, \sin t, t)$.
This is the equation of a *helix* (see Figure 1.8.1). As the value of t increases, the terminal points of $\mathbf{f}(t)$ trace out a curve spiraling upward. For each t, the x- and y-coordinates of $\mathbf{f}(t)$ are $x = \cos t$ and $y = \sin t$, so

$$x^2 + y^2 = \cos^2 t + \sin^2 t = 1.$$

Thus, the curve lies on the surface of the right circular cylinder $x^2 + y^2 = 1$.

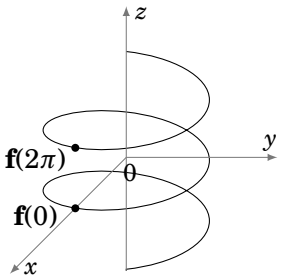


Figure 1.8.1

It may help to think of vector-valued functions of a real variable in $\mathbb{R}^3$ as a generalization of the parametric functions in $\mathbb{R}^2$ which you learned about in single-variable calculus. Much of the theory of real-valued functions of a single real variable can be applied to vector-valued functions of a real variable. Since each of the three component functions are real-valued, it will sometimes be the case that results from single-variable calculus can simply be applied to each of the component functions to yield a similar result for the vector-valued function. However, there are times when such generalizations do not hold (see Exercise 13). The concept of a limit, though, can be extended naturally to vector-valued functions, as in the following definition.

Definition 1.11. Let $\mathbf{f}(t)$ be a vector-valued function, let a be a real number and let $\mathbf{c}$ be a vector. Then we say that the **limit** of $\mathbf{f}(t)$ as t approaches a equals $\mathbf{c}$, written as $\lim_{t\to a}\mathbf{f}(t)=\mathbf{c}$, if $\lim_{t\to a}\|\mathbf{f}(t)-\mathbf{c}\|=0$. If $\mathbf{f}(t)=(f_1(t),f_2(t),f_3(t))$, then

$$\lim_{t\to a}\mathbf{f}(t)=\left(\lim_{t\to a}f_1(t),\lim_{t\to a}f_2(t),\lim_{t\to a}f_3(t)\right)$$

provided that all three limits on the right side exist.

The above definition shows that continuity and the derivative of vector-valued functions can also be defined in terms of its component functions.

Definition 1.12. Let $\mathbf{f}(t)=(f_1(t),f_2(t),f_3(t))$ be a vector-valued function, and let a be a real number in its domain. Then $\mathbf{f}(t)$ is **continuous** at a if $\lim_{t\to a}\mathbf{f}(t)=\mathbf{f}(a)$. Equivalently, $\mathbf{f}(t)$ is continuous at a if and only if $f_1(t)$, $f_2(t)$, and $f_3(t)$ are continuous at a.
The **derivative** of $\mathbf{f}(t)$ at a, denoted by $\mathbf{f}'(a)$ or $\dfrac{d\mathbf{f}}{dt}(a)$, is the limit

$$\mathbf{f}'(a)=\lim_{h\to 0}\frac{\mathbf{f}(a+h)-\mathbf{f}(a)}{h}$$

if that limit exists. Equivalently, $\mathbf{f}'(a)=(f_1{}'(a),f_2{}'(a),f_3{}'(a))$, if the component derivatives exist. We say that $\mathbf{f}(t)$ is **differentiable** at a if $\mathbf{f}'(a)$ exists.

Recall that the derivative of a real-valued function of a single variable is a real number, representing the slope of the tangent line to the graph of the function at a point. Similarly, the derivative of a vector-valued function is a **tangent vector** to the curve in space which the function represents, and it lies on the *tangent line* to the curve (see Figure 1.8.2).

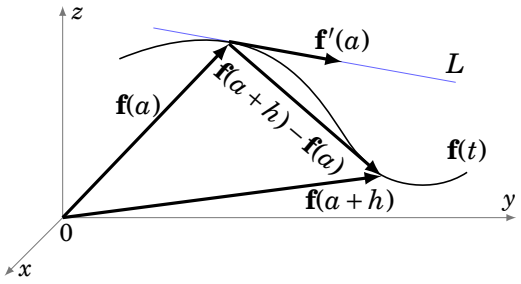


Figure 1.8.2 Tangent vector $\mathbf{f}'(a)$ and tangent line $L=\mathbf{f}(a)+s\mathbf{f}'(a)$

Example 1.36. Let $\mathbf{f}(t)=(\cos t,\sin t,t)$. Then $\mathbf{f}'(t)=(-\sin t,\cos t,1)$ for all t. The tangent line L to the curve at $\mathbf{f}(2\pi)=(1,0,2\pi)$ is $L=\mathbf{f}(2\pi)+s\,\mathbf{f}'(2\pi)=(1,0,2\pi)+s(0,1,1)$, or in parametric form: $x=1$, $y=s$, $z=2\pi+s$ for $-\infty<s<\infty$.

A **scalar function** is a real-valued function. Note that if $u(t)$ is a scalar function and $\mathbf{f}(t)$ is a vector-valued function, then their product, defined by $(u\,\mathbf{f})(t) = u(t)\mathbf{f}(t)$ for all t, is a vector-valued function (since the product of a scalar with a vector is a vector).

The basic properties of derivatives of vector-valued functions are summarized in the following theorem.

Theorem 1.20. Let $\mathbf{f}(t)$ and $\mathbf{g}(t)$ be differentiable vector-valued functions, let $u(t)$ be a differentiable scalar function, let k be a scalar, and let $\mathbf{c}$ be a constant vector. Then

(a) $\dfrac{d}{dt}(\mathbf{c}) = \mathbf{0}$

(b) $\dfrac{d}{dt}(k\mathbf{f}) = k\dfrac{d\mathbf{f}}{dt}$

(c) $\dfrac{d}{dt}(\mathbf{f}+\mathbf{g}) = \dfrac{d\mathbf{f}}{dt} + \dfrac{d\mathbf{g}}{dt}$

(d) $\dfrac{d}{dt}(\mathbf{f}-\mathbf{g}) = \dfrac{d\mathbf{f}}{dt} - \dfrac{d\mathbf{g}}{dt}$

(e) $\dfrac{d}{dt}(u\,\mathbf{f}) = \dfrac{du}{dt}\mathbf{f} + u\dfrac{d\mathbf{f}}{dt}$

(f) $\dfrac{d}{dt}(\mathbf{f}\cdot\mathbf{g}) = \dfrac{d\mathbf{f}}{dt}\cdot\mathbf{g} + \mathbf{f}\cdot\dfrac{d\mathbf{g}}{dt}$

(g) $\dfrac{d}{dt}(\mathbf{f}\times\mathbf{g}) = \dfrac{d\mathbf{f}}{dt}\times\mathbf{g} + \mathbf{f}\times\dfrac{d\mathbf{g}}{dt}$

Proof: The proofs of parts (a)-(e) follow easily by differentiating the component functions and using the rules for derivatives from single-variable calculus. We will prove part (f), and leave the proof of part (g) as an exercise for the reader.

(f) Write $\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$ and $\mathbf{g}(t) = (g_1(t), g_2(t), g_3(t))$, where the component functions $f_1(t)$, $f_2(t)$, $f_3(t)$, $g_1(t)$, $g_2(t)$, $g_3(t)$ are all differentiable real-valued functions. Then

$$\begin{aligned}
\frac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{g}(t)) &= \frac{d}{dt}(f_1(t)g_1(t) + f_2(t)g_2(t) + f_3(t)g_3(t)) \\
&= \frac{d}{dt}(f_1(t)g_1(t)) + \frac{d}{dt}(f_2(t)g_2(t)) + \frac{d}{dt}(f_3(t)g_3(t)) \\
&= \frac{df_1}{dt}(t)g_1(t) + f_1(t)\frac{dg_1}{dt}(t) + \frac{df_2}{dt}(t)g_2(t) + f_2(t)\frac{dg_2}{dt}(t) + \frac{df_3}{dt}(t)g_3(t) + f_3(t)\frac{dg_3}{dt}(t) \\
&= \left(\frac{df_1}{dt}(t), \frac{df_2}{dt}(t), \frac{df_3}{dt}(t)\right)\cdot(g_1(t), g_2(t), g_3(t)) \\
&\quad + (f_1(t), f_2(t), f_3(t))\cdot\left(\frac{dg_1}{dt}(t), \frac{dg_2}{dt}(t), \frac{dg_3}{dt}(t)\right) \\
&= \frac{d\mathbf{f}}{dt}(t)\cdot\mathbf{g}(t) + \mathbf{f}(t)\cdot\frac{d\mathbf{g}}{dt}(t) \quad \text{for all } t.
\end{aligned}$$

QED

Example 1.37. Suppose $\mathbf{f}(t)$ is differentiable. Find the derivative of $\|\mathbf{f}(t)\|$.

Solution: Since $\|\mathbf{f}(t)\|$ is a real-valued function of t, then by the Chain Rule for real-valued functions, we know that $\dfrac{d}{dt}\|\mathbf{f}(t)\|^2 = 2\|\mathbf{f}(t)\|\dfrac{d}{dt}\|\mathbf{f}(t)\|$.

But $\|\mathbf{f}(t)\|^2 = \mathbf{f}(t)\cdot\mathbf{f}(t)$, so $\dfrac{d}{dt}\|\mathbf{f}(t)\|^2 = \dfrac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{f}(t))$. Hence, we have

$$\begin{aligned} 2\|\mathbf{f}(t)\|\frac{d}{dt}\|\mathbf{f}(t)\| &= \frac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{f}(t)) = \mathbf{f}'(t)\cdot\mathbf{f}(t) + \mathbf{f}(t)\cdot\mathbf{f}'(t) \text{ by Theorem 1.20(f), so} \\ &= 2\mathbf{f}'(t)\cdot\mathbf{f}(t) \text{ , so if } \|\mathbf{f}(t)\| \neq 0 \text{ then} \\ \frac{d}{dt}\|\mathbf{f}(t)\| &= \frac{\mathbf{f}'(t)\cdot\mathbf{f}(t)}{\|\mathbf{f}(t)\|}. \end{aligned}$$

We know that $\|\mathbf{f}(t)\|$ is constant if and only if $\dfrac{d}{dt}\|\mathbf{f}(t)\| = 0$ for all t. Also, $\mathbf{f}(t) \perp \mathbf{f}'(t)$ if and only if $\mathbf{f}'(t)\cdot\mathbf{f}(t) = 0$. Thus, the above example shows this important fact:

If $\|\mathbf{f}(t)\| \neq 0$, then $\|\mathbf{f}(t)\|$ is constant if and only if $\mathbf{f}(t) \perp \mathbf{f}'(t)$ for all t.

This means that if a curve lies completely on a sphere (or circle) centered at the origin, then the tangent vector $\mathbf{f}'(t)$ is always perpendicular to the *position vector* $\mathbf{f}(t)$.

Example 1.38. The *spherical spiral* $\mathbf{f}(t) = \left(\dfrac{\cos t}{\sqrt{1+a^2t^2}}, \dfrac{\sin t}{\sqrt{1+a^2t^2}}, \dfrac{-at}{\sqrt{1+a^2t^2}}\right)$, for $a \neq 0$.

Figure 1.8.3 shows the graph of the curve when $a = 0.2$. In the exercises, the reader will be asked to show that this curve lies on the sphere $x^2 + y^2 + z^2 = 1$ and to verify directly that $\mathbf{f}'(t)\cdot\mathbf{f}(t) = 0$ for all t.

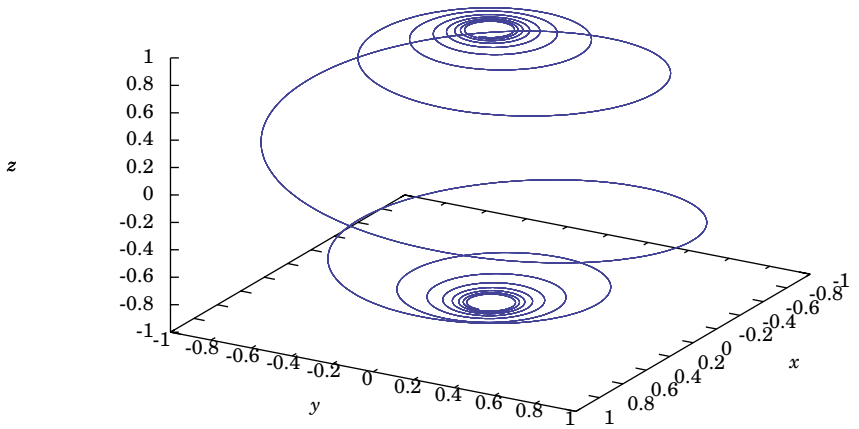


Figure 1.8.3 Spherical spiral with $a = 0.2$

Just as in single-variable calculus, higher-order derivatives of vector-valued functions are obtained by repeatedly differentiating the (first) derivative of the function:

$$\mathbf{f}''(t) = \frac{d}{dt}\mathbf{f}'(t), \qquad \mathbf{f}'''(t) = \frac{d}{dt}\mathbf{f}''(t), \quad \ldots \quad , \quad \frac{d^n\mathbf{f}}{dt^n} = \frac{d}{dt}\left(\frac{d^{n-1}\mathbf{f}}{dt^{n-1}}\right) \quad (\text{for } n = 2,3,4,\ldots)$$

We can use vector-valued functions to represent physical quantities, such as velocity, acceleration, force, momentum, etc. For example, let the real variable t represent time elapsed from some initial time ($t = 0$), and suppose that an object of constant mass m is subjected to some force so that it moves in space, with its position (x, y, z) at time t a function of t. That is, $x = x(t)$, $y = y(t)$, $z = z(t)$ for some real-valued functions $x(t)$, $y(t)$, $z(t)$. Call $\mathbf{r}(t) = (x(t), y(t), z(t))$ the **position vector** of the object. We can define various physical quantities associated with the object as follows:[14]

$$\begin{aligned}
\textit{position}\colon\ & \mathbf{r}(t) = (x(t), y(t), z(t)) \\
\textit{velocity}\colon\ & \mathbf{v}(t) = \dot{\mathbf{r}}(t) = \mathbf{r}'(t) = \frac{d\mathbf{r}}{dt} \\
& = (x'(t), y'(t), z'(t)) \\
\textit{acceleration}\colon\ & \mathbf{a}(t) = \dot{\mathbf{v}}(t) = \mathbf{v}'(t) = \frac{d\mathbf{v}}{dt} \\
& = \ddot{\mathbf{r}}(t) = \mathbf{r}''(t) = \frac{d^2\mathbf{r}}{dt^2} \\
& = (x''(t), y''(t), z''(t)) \\
\textit{momentum}\colon\ & \mathbf{p}(t) = m\mathbf{v}(t) \\
\textit{force}\colon\ & \mathbf{F}(t) = \dot{\mathbf{p}}(t) = \mathbf{p}'(t) = \frac{d\mathbf{p}}{dt} \quad \text{(Newton's Second Law of Motion)}
\end{aligned}$$

The magnitude $\|\mathbf{v}(t)\|$ of the velocity vector is called the *speed* of the object. Note that since the mass m is a constant, the force equation becomes the familiar $\mathbf{F}(t) = m\mathbf{a}(t)$.

Example 1.39. Let $\mathbf{r}(t) = (5\cos t, 3\sin t, 4\sin t)$ be the position vector of an object at time $t \geq 0$. Find its (a) velocity and (b) acceleration vectors.

Solution: (a) $\mathbf{v}(t) = \dot{\mathbf{r}}(t) = (-5\sin t, 3\cos t, 4\cos t)$

(b) $\mathbf{a}(t) = \dot{\mathbf{v}}(t) = (-5\cos t, -3\sin t, -4\sin t)$

Note that $\|\mathbf{r}(t)\| = \sqrt{25\cos^2 t + 25\sin^2 t} = 5$ for all t, so by Example 1.37 we know that $\mathbf{r}(t) \cdot \dot{\mathbf{r}}(t) = 0$ for all t (which we can verify from part (a)). In fact, $\|\mathbf{v}(t)\| = 5$ for all t also. And not only does $\mathbf{r}(t)$ lie on the sphere of radius 5 centered at the origin, but perhaps not so obvious is that it lies completely within a *circle* of radius 5 centered at the origin. Also, note that $\mathbf{a}(t) = -\mathbf{r}(t)$. It turns out (see Exercise 16) that whenever an object moves in a circle with constant speed, the acceleration vector will point in the opposite direction of the position vector (i.e. towards the center of the circle).

---

[14] We will often use the older dot notation for derivatives when physics is involved.

Recall from Section 1.5 that if $\mathbf{r}_1$, $\mathbf{r}_2$ are position vectors to distinct points then $\mathbf{r}_1 + t(\mathbf{r}_2 - \mathbf{r}_1)$ represents a line through those two points as t varies over all real numbers. That vector sum can be written as $(1-t)\mathbf{r}_1 + t\mathbf{r}_2$. So the function $\mathbf{l}(t) = (1-t)\mathbf{r}_1 + t\mathbf{r}_2$ is a line through the terminal points of $\mathbf{r}_1$ and $\mathbf{r}_2$, and when t is restricted to the interval $[0,1]$ it is the line segment between the points, with $\mathbf{l}(0) = \mathbf{r}_1$ and $\mathbf{l}(1) = \mathbf{r}_2$.

In general, a function of the form $\mathbf{f}(t) = (a_1 t + b_1, a_2 t + b_2, a_3 t + b_3)$ represents a line in $\mathbb{R}^3$. A function of the form $\mathbf{f}(t) = (a_1 t^2 + b_1 t + c_1, a_2 t^2 + b_2 t + c_2, a_3 t^2 + b_3 t + c_3)$ represents a (possibly degenerate) parabola in $\mathbb{R}^3$.

Example 1.40. *Bézier curves* are used in Computer Aided Design (CAD) to approximate the shape of a polygonal path in space (called the *Bézier polygon* or *control polygon*). For instance, given three points (or position vectors) $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$ in $\mathbb{R}^3$, define

$$\begin{aligned}
\mathbf{b}_0^1(t) &= (1-t)\mathbf{b}_0 + t\mathbf{b}_1 \\
\mathbf{b}_1^1(t) &= (1-t)\mathbf{b}_1 + t\mathbf{b}_2 \\
\mathbf{b}_0^2(t) &= (1-t)\mathbf{b}_0^1(t) + t\mathbf{b}_1^1(t) \\
&= (1-t)^2\mathbf{b}_0 + 2t(1-t)\mathbf{b}_1 + t^2\mathbf{b}_2
\end{aligned}$$

for all real t. For t in the interval $[0,1]$, we see that $\mathbf{b}_0^1(t)$ is the line segment between $\mathbf{b}_0$ and $\mathbf{b}_1$, and $\mathbf{b}_1^1(t)$ is the line segment between $\mathbf{b}_1$ and $\mathbf{b}_2$. The function $\mathbf{b}_0^2(t)$ is the Bézier curve for the points $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$. Note from the last formula that the curve is a parabola that goes through $\mathbf{b}_0$ (when $t = 0$) and $\mathbf{b}_2$ (when $t = 1$).

As an example, let $\mathbf{b}_0 = (0,0,0)$, $\mathbf{b}_1 = (1,2,3)$, and $\mathbf{b}_2 = (4,5,2)$. Then the explicit formula for the Bézier curve is $\mathbf{b}_0^2(t) = (2t + 2t^2, 4t + t^2, 6t - 4t^2)$, as shown in Figure 1.8.4, where the line segments are $\mathbf{b}_0^1(t)$ and $\mathbf{b}_1^1(t)$, and the curve is $\mathbf{b}_0^2(t)$.

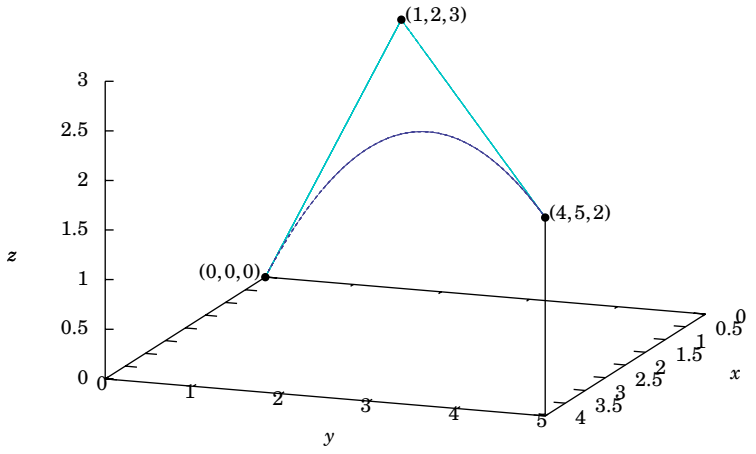


Figure 1.8.4 Bézier curve approximation for three points

In general, the polygonal path determined by $n \geq 3$ noncollinear points in $\mathbb{R}^3$ can be used to define the Bézier curve recursively by a process called *repeated linear interpolation*. This curve will be a vector-valued function whose components are polynomials of degree $n-1$, and its formula is given by *de Casteljau's algorithm*.[15] In the exercises, the reader will be given the algorithm for the case of $n = 4$ points and asked to write the explicit formula for the Bézier curve for the four points shown in Figure 1.8.5.

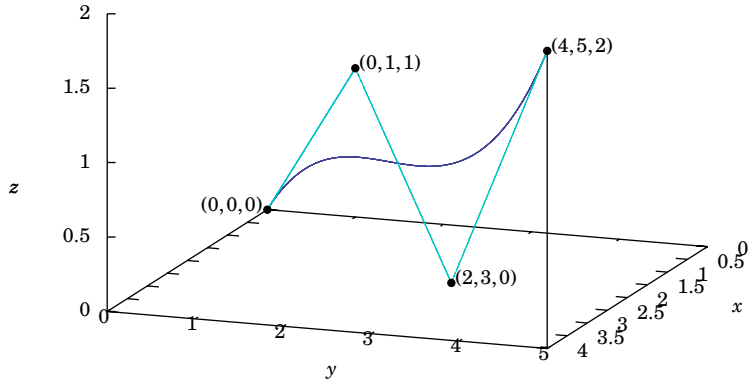


Figure 1.8.5 Bézier curve approximation for four points

Exercises

A

For Exercises 1-4, calculate $\mathbf{f}'(t)$ and find the tangent line at $\mathbf{f}(0)$.

1. $\mathbf{f}(t) = (t+1, t^2+1, t^3+1)$

2. $\mathbf{f}(t) = (e^t+1, e^{2t}+1, e^{t^2}+1)$

3. $\mathbf{f}(t) = (\cos 2t, \sin 2t, t)$

4. $\mathbf{f}(t) = (\sin 2t, 2\sin^2 t, 2\cos t)$

For Exercises 5-6, find the velocity $\mathbf{v}(t)$ and acceleration $\mathbf{a}(t)$ of an object with the given position vector $\mathbf{r}(t)$.

5. $\mathbf{r}(t) = (t, t-\sin t, 1-\cos t)$

6. $\mathbf{r}(t) = (3\cos t, 2\sin t, 1)$

B

7. Let $\mathbf{f}(t) = \left(\dfrac{\cos t}{\sqrt{1+a^2t^2}}, \dfrac{\sin t}{\sqrt{1+a^2t^2}}, \dfrac{-at}{\sqrt{1+a^2t^2}}\right)$, with $a \neq 0$.

(a) Show that $\|\mathbf{f}(t)\| = 1$ for all t.

(b) Show directly that $\mathbf{f}'(t) \cdot \mathbf{f}(t) = 0$ for all t.

8. If $\mathbf{f}'(t) = \mathbf{0}$ for all t in some interval (a, b), show that $\mathbf{f}(t)$ is a constant vector in (a, b).

[15] See pp. 27-30 in FARIN.

9. For a constant vector $\mathbf{c} \neq \mathbf{0}$, the function $\mathbf{f}(t) = t\mathbf{c}$ represents a line parallel to $\mathbf{c}$.

(a) What kind of curve does $\mathbf{g}(t) = t^3\mathbf{c}$ represent? Explain.

(b) What kind of curve does $\mathbf{h}(t) = e^t\mathbf{c}$ represent? Explain.

(c) Compare $\mathbf{f}'(0)$ and $\mathbf{g}'(0)$. Given your answer to part (a), how do you explain the difference in the two derivatives?

10. Show that $\dfrac{d}{dt}\left(\mathbf{f} \times \dfrac{d\mathbf{f}}{dt}\right) = \mathbf{f} \times \dfrac{d^2\mathbf{f}}{dt^2}$.

11. Let a particle of (constant) mass m have position vector $\mathbf{r}(t)$, velocity $\mathbf{v}(t)$, acceleration $\mathbf{a}(t)$ and momentum $\mathbf{p}(t)$ at time t. The *angular momentum* $\mathbf{L}(t)$ of the particle with respect to the origin at time t is defined as $\mathbf{L}(t) = \mathbf{r}(t) \times \mathbf{p}(t)$. If $\mathbf{F}(t)$ is the force acting on the particle at time t, then define the *torque* $\mathbf{N}(t)$ acting on the particle with respect to the origin as $\mathbf{N}(t) = \mathbf{r}(t) \times \mathbf{F}(t)$. Show that $\mathbf{L}'(t) = \mathbf{N}(t)$.

12. Show that $\dfrac{d}{dt}(\mathbf{f} \cdot (\mathbf{g} \times \mathbf{h})) = \dfrac{d\mathbf{f}}{dt} \cdot (\mathbf{g} \times \mathbf{h}) + \mathbf{f} \cdot \left(\dfrac{d\mathbf{g}}{dt} \times \mathbf{h}\right) + \mathbf{f} \cdot \left(\mathbf{g} \times \dfrac{d\mathbf{h}}{dt}\right)$.

13. The Mean Value Theorem does not hold for vector-valued functions: Show that for $\mathbf{f}(t) = (\cos t, \sin t, t)$, there is no t in the interval $(0, 2\pi)$ such that

$$\mathbf{f}'(t) = \frac{\mathbf{f}(2\pi) - \mathbf{f}(0)}{2\pi - 0}.$$

C

14. The Bézier curve $\mathbf{b}_0^3(t)$ for four noncollinear points $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ in $\mathbb{R}^3$ is defined by the following algorithm (going from the left column to the right):

$$\mathbf{b}_0^1(t) = (1-t)\mathbf{b}_0 + t\mathbf{b}_1 \qquad \mathbf{b}_0^2(t) = (1-t)\mathbf{b}_0^1(t) + t\mathbf{b}_1^1(t) \qquad \mathbf{b}_0^3(t) = (1-t)\mathbf{b}_0^2(t) + t\mathbf{b}_1^2(t)$$
$$\mathbf{b}_1^1(t) = (1-t)\mathbf{b}_1 + t\mathbf{b}_2 \qquad \mathbf{b}_1^2(t) = (1-t)\mathbf{b}_1^1(t) + t\mathbf{b}_2^1(t)$$
$$\mathbf{b}_2^1(t) = (1-t)\mathbf{b}_2 + t\mathbf{b}_3$$

(a) Show that $\mathbf{b}_0^3(t) = (1-t)^3\mathbf{b}_0 + 3t(1-t)^2\mathbf{b}_1 + 3t^2(1-t)\mathbf{b}_2 + t^3\mathbf{b}_3$.

(b) Write the explicit formula (as in Example 1.40) for the Bézier curve for the points $\mathbf{b}_0 = (0,0,0)$, $\mathbf{b}_1 = (0,1,1)$, $\mathbf{b}_2 = (2,3,0)$, $\mathbf{b}_3 = (4,5,2)$.

15. Let $\mathbf{r}(t)$ be the position vector for a particle moving in $\mathbb{R}^3$. Show that

$$\frac{d}{dt}(\mathbf{r} \times (\mathbf{v} \times \mathbf{r})) = \|\mathbf{r}\|^2\mathbf{a} + (\mathbf{r} \cdot \mathbf{v})\mathbf{v} - (\|\mathbf{v}\|^2 + \mathbf{r} \cdot \mathbf{a})\mathbf{r}.$$

16. Let $\mathbf{r}(t)$ be the position vector in $\mathbb{R}^3$ for a particle that moves with constant speed $c > 0$ in a circle of radius $a > 0$ in the xy-plane. Show that $\mathbf{a}(t)$ points in the opposite direction as $\mathbf{r}(t)$ for all t. (*Hint: Use Example 1.37 to show that* $\mathbf{r}(t) \perp \mathbf{v}(t)$ *and* $\mathbf{a}(t) \perp \mathbf{v}(t)$, *and hence* $\mathbf{a}(t) \parallel \mathbf{r}(t)$.)

17. Prove Theorem 1.20(g).

1.9 Arc Length

Let $\mathbf{r}(t) = (x(t), y(t), z(t))$ be the position vector of an object moving in $\mathbb{R}^3$. Since $\|\mathbf{v}(t)\|$ is the speed of the object at time t, it seems natural to define the distance s traveled by the object from time $t = a$ to $t = b$ as the definite integral

$$s = \int_a^b \|\mathbf{v}(t)\|\, dt = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\, dt, \tag{1.40}$$

which is analogous to the case from single-variable calculus for parametric functions in $\mathbb{R}^2$. This is indeed how we will define the distance traveled and, in general, the arc length of a curve in $\mathbb{R}^3$.

Definition 1.13. Let $\mathbf{f}(t) = (x(t), y(t), z(t))$ be a curve in $\mathbb{R}^3$ whose domain includes the interval $[a, b]$. Suppose that in the interval (a, b) the first derivative of each component function $x(t)$, $y(t)$ and $z(t)$ exists and is continuous, and that no section of the curve is repeated. Then the **arc length** L of the curve from $t = a$ to $t = b$ is

$$L = \int_a^b \|\mathbf{f}'(t)\|\, dt = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\, dt \tag{1.41}$$

A real-valued function whose first derivative is continuous is called *continuously differentiable* (or a $\mathscr{C}^1$ function), and a function whose derivatives of all orders are continuous is called *smooth* (or a $\mathscr{C}^\infty$ function). All the functions we will consider will be smooth. A *smooth curve* $\mathbf{f}(t)$ is one whose derivative $\mathbf{f}'(t)$ is never the zero vector and whose component functions are all smooth.

Note that we did not *prove* that the formula in the above definition actually gives the length of a section of a curve. A rigorous proof requires dealing with some subtleties, normally glossed over in calculus texts, which are beyond the scope of this book.[16]

Example 1.41. Find the length L of the helix $\mathbf{f}(t) = (\cos t, \sin t, t)$ from $t = 0$ to $t = 2\pi$.

Solution: By formula (1.41), we have

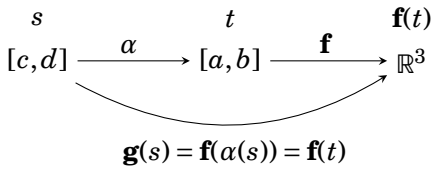
$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{(-\sin t)^2 + (\cos t)^2 + 1^2}\, dt = \int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t + 1}\, dt = \int_0^{2\pi} \sqrt{2}\, dt \\ &= \sqrt{2}(2\pi - 0) = 2\sqrt{2}\pi \end{aligned}$$

Similar to the case in $\mathbb{R}^2$, if there are values of t in the interval $[a, b]$ where the derivative of a component function is not continuous then it is often possible to partition $[a, b]$ into subintervals where all the component functions are continuously differentiable (except at the endpoints, which can be ignored). The sum of the arc lengths over the subintervals will be the arc length over $[a, b]$.

[16] In particular, *Duhamel's principle* is needed. See the proof in TAYLOR and MANN, § 14.2 and § 18.2.

Notice that the curve traced out by the function $\mathbf{f}(t) = (\cos t, \sin t, t)$ from Example 1.41 is also traced out by the function $\mathbf{g}(t) = (\cos 2t, \sin 2t, 2t)$. For example, over the interval $[0, \pi]$, $\mathbf{g}(t)$ traces out the same section of the curve as $\mathbf{f}(t)$ does over the interval $[0, 2\pi]$. Intuitively, this says that $\mathbf{g}(t)$ traces the curve twice as fast as $\mathbf{f}(t)$. This makes sense since, viewing the functions as position vectors and their derivatives as velocity vectors, the speeds of $\mathbf{f}(t)$ and $\mathbf{g}(t)$ are $\|\mathbf{f}'(t)\| = \sqrt{2}$ and $\|\mathbf{g}'(t)\| = 2\sqrt{2}$, respectively. We say that $\mathbf{g}(t)$ and $\mathbf{f}(t)$ are different *parametrizations* of the same curve.

Definition 1.14. Let C be a smooth curve in $\mathbb{R}^3$ represented by a function $\mathbf{f}(t)$ defined on an interval $[a, b]$, and let $\alpha : [c, d] \to [a, b]$ be a smooth one-to-one mapping of an interval $[c, d]$ onto $[a, b]$. Then the function $\mathbf{g} : [c, d] \to \mathbb{R}^3$ defined by $\mathbf{g}(s) = \mathbf{f}(\alpha(s))$ is a **parametrization** of C with **parameter** s. If α is strictly increasing on $[c, d]$ then we say that $\mathbf{g}(s)$ is *equivalent* to $\mathbf{f}(t)$.



Note that the differentiability of $\mathbf{g}(s)$ follows from a version of the Chain Rule for vector-valued functions (the proof is left as an exercise):

Theorem 1.21. Chain Rule: If $\mathbf{f}(t)$ is a differentiable vector-valued function of t, and $t = \alpha(s)$ is a differentiable scalar function of s, then $\mathbf{f}(s) = \mathbf{f}(\alpha(s))$ is a differentiable vector-valued function of s, and

$$\frac{d\mathbf{f}}{ds} = \frac{d\mathbf{f}}{dt}\frac{dt}{ds} \tag{1.42}$$

for any s where the composite function $\mathbf{f}(\alpha(s))$ is defined.

Example 1.42. The following are all equivalent parametrizations of the same curve:

$$\mathbf{f}(t) = (\cos t, \sin t, t) \text{ for } t \text{ in } [0, 2\pi]$$
$$\mathbf{g}(s) = (\cos 2s, \sin 2s, 2s) \text{ for } s \text{ in } [0, \pi]$$
$$\mathbf{h}(s) = (\cos 2\pi s, \sin 2\pi s, 2\pi s) \text{ for } s \text{ in } [0, 1]$$

To see that $\mathbf{g}(s)$ is equivalent to $\mathbf{f}(t)$, define $\alpha : [0, \pi] \to [0, 2\pi]$ by $\alpha(s) = 2s$. Then α is smooth, one-to-one, maps $[0, \pi]$ onto $[0, 2\pi]$, and is strictly increasing (since $\alpha'(s) = 2 > 0$ for all s). Likewise, defining $\alpha : [0, 1] \to [0, 2\pi]$ by $\alpha(s) = 2\pi s$ shows that $\mathbf{h}(s)$ is equivalent to $\mathbf{f}(t)$.

A curve can have many parametrizations, with different speeds, so which one is the best to use? In some situations the **arc length parametrization** can be useful. The idea behind this is to replace the parameter t, for any given smooth parametrization $\mathbf{f}(t)$ defined on $[a,b]$, by the parameter s given by

$$s = s(t) = \int_a^t \|\mathbf{f}'(u)\|\,du. \tag{1.43}$$

In terms of motion along a curve, s is the distance traveled along the curve after time t has elapsed. So the new parameter will be distance instead of time. There is a natural correspondence between s and t: from a starting point on the curve, the distance traveled along the curve (in one direction) is uniquely determined by the amount of time elapsed, and vice versa.

Since s is the arc length of the curve over the interval $[a,t]$ for each t in $[a,b]$, then it is a function of t. By the Fundamental Theorem of Calculus, its derivative is

$$s'(t) = \frac{ds}{dt} = \frac{d}{dt}\int_a^t \|\mathbf{f}'(u)\|\,du = \|\mathbf{f}'(t)\| \text{ for all } t \text{ in } [a,b].$$

Since $\mathbf{f}(t)$ is smooth, then $\|\mathbf{f}'(t)\| > 0$ for all t in $[a,b]$. Thus $s'(t) > 0$ and hence $s(t)$ is strictly increasing on the interval $[a,b]$. Recall that this means that s is a one-to-one mapping of the interval $[a,b]$ onto the interval $[s(a),s(b)]$. But we see that

$$s(a) = \int_a^a \|\mathbf{f}'(u)\|\,du = 0 \quad \text{and} \quad s(b) = \int_a^b \|\mathbf{f}'(u)\|\,du = L = \text{arc length from } t = a \text{ to } t = b$$

So the function $s : [a,b] \to [0,L]$ is a one-to-one, differentiable mapping onto the interval $[0,L]$. From single-variable calculus, we know that this means that there exists an inverse function $\alpha : [0,L] \to [a,b]$ that is differentiable and the inverse of $s : [a,b] \to [0,L]$. That is, for each t in $[a,b]$ there is a unique s in $[0,L]$ such that $s = s(t)$ and $t = \alpha(s)$. And we know that the derivative of α is

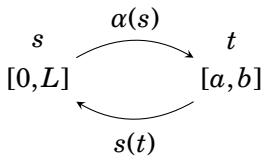


Figure 1.9.1 $t = \alpha(s)$

$$\alpha'(s) = \frac{1}{s'(\alpha(s))} = \frac{1}{\|\mathbf{f}'(\alpha(s))\|}$$

So define the arc length parametrization $\mathbf{f} : [0,L] \to \mathbb{R}^3$ by

$$\mathbf{f}(s) = \mathbf{f}(\alpha(s)) \text{ for all } s \text{ in } [0,L].$$

Then $\mathbf{f}(s)$ is smooth, by the Chain Rule. In fact, $\mathbf{f}(s)$ has *unit speed*:

$$\begin{aligned} \mathbf{f}'(s) &= \mathbf{f}'(\alpha(s))\,\alpha'(s) \text{ by the Chain Rule, so} \\ &= \mathbf{f}'(\alpha(s))\,\frac{1}{\|\mathbf{f}'(\alpha(s))\|} \text{ , so} \\ \|\mathbf{f}'(s)\| &= 1 \text{ for all } s \text{ in } [0,L]. \end{aligned}$$

So the arc length parametrization traverses the curve at a "normal" rate.

In practice, parametrizing a curve $\mathbf{f}(t)$ by arc length requires you to evaluate the integral $s = \int_a^t \|\mathbf{f}'(u)\|\,du$ in some closed form (as a function of t) so that you could then solve for t in terms of s. If that can be done, you would then substitute the expression for t in terms of s (which we called $\alpha(s)$) into the formula for $\mathbf{f}(t)$ to get $\mathbf{f}(s)$.

Example 1.43. Parametrize the helix $\mathbf{f}(t) = (\cos t, \sin t, t)$, for t in $[0, 2\pi]$, by arc length.

Solution: By Example 1.41 and formula (1.43), we have

$$s = \int_0^t \|\mathbf{f}'(u)\|\,du = \int_0^t \sqrt{2}\,du = \sqrt{2}\,t \quad \text{for all } t \text{ in } [0, 2\pi].$$

So we can solve for t in terms of s: $t = \alpha(s) = \dfrac{s}{\sqrt{2}}$.

$\therefore\ \mathbf{f}(s) = \left(\cos\dfrac{s}{\sqrt{2}}, \sin\dfrac{s}{\sqrt{2}}, \dfrac{s}{\sqrt{2}}\right)$ for all s in $[0, 2\sqrt{2}\pi]$. Note that $\|\mathbf{f}'(s)\| = 1$.

Arc length plays an important role when discussing *curvature* and *moving frame fields*, in the field of mathematics known as *differential geometry*.[17] The methods involve using an arc length parametrization, which often leads to an integral that is either difficult or impossible to evaluate in a simple closed form. The simple integral in Example 1.43 is the exception, not the norm. In general, arc length parametrizations are more useful for theoretical purposes than for practical computations.[18] Curvature and moving frame fields can be defined without using arc length, which makes their computation much easier, and these definitions can be shown to be equivalent to those using arc length. We will leave this to the exercises.

The arc length for curves given in other coordinate systems can also be calculated:

Theorem 1.22. Suppose that $r = r(t)$, $\theta = \theta(t)$ and $z = z(t)$ are the cylindrical coordinates of a curve $\mathbf{f}(t)$, for t in $[a, b]$. Then the arc length L of the curve over $[a, b]$ is

$$L = \int_a^b \sqrt{r'(t)^2 + r(t)^2\theta'(t)^2 + z'(t)^2}\,dt \tag{1.44}$$

Proof: The Cartesian coordinates $(x(t), y(t), z(t))$ of a point on the curve are given by

$$x(t) = r(t)\cos\theta(t), \quad y(t) = r(t)\sin\theta(t), \quad z(t) = z(t)$$

so differentiating the above expressions for $x(t)$ and $y(t)$ with respect to t gives

$$x'(t) = r'(t)\cos\theta(t) - r(t)\theta'(t)\sin\theta(t), \quad y'(t) = r'(t)\sin\theta(t) + r(t)\theta'(t)\cos\theta(t)$$

[17] See O'NEILL for an introduction to elementary differential geometry.

[18] For example, the usual parametrizations of Bézier curves, which we discussed in Section 1.8, are polynomial functions in $\mathbb{R}^3$. This makes their computation relatively simple, which, in CAD, is desirable. But their arc length parametrizations are not only *not* polynomials, they are in fact usually impossible to calculate at all.

and so

$$\begin{aligned} x'(t)^2 + y'(t)^2 &= (r'(t)\cos\theta(t) - r(t)\theta'(t)\sin\theta(t))^2 + (r'(t)\sin\theta(t) + r(t)\theta'(t)\cos\theta(t))^2 \\ &= r'(t)^2(\cos^2\theta + \sin^2\theta) + r(t)^2\theta'(t)^2(\cos^2\theta + \sin^2\theta) \\ &\quad - 2r'(t)r(t)\theta'(t)\cos\theta\sin\theta + 2r'(t)r(t)\theta'(t)\cos\theta\sin\theta \\ &= r'(t)^2 + r(t)^2\theta'(t)^2 \quad \text{, and so} \\ L &= \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\, dt \\ &= \int_a^b \sqrt{r'(t)^2 + r(t)^2\theta'(t)^2 + z'(t)^2}\, dt \end{aligned}$$

QED

Example 1.44. Find the arc length L of the curve whose cylindrical coordinates are $r = e^t$, $\theta = t$ and $z = e^t$, for t over the interval $[0,1]$.

Solution: Since $r'(t) = e^t$, $\theta'(t) = 1$ and $z'(t) = e^t$, then

$$\begin{aligned} L &= \int_0^1 \sqrt{r'(t)^2 + r(t)^2\theta'(t)^2 + z'(t)^2}\, dt \\ &= \int_0^1 \sqrt{e^{2t} + e^{2t}(1) + e^{2t}}\, dt \\ &= \int_0^1 e^t\sqrt{3}\, dt = \sqrt{3}(e-1) \end{aligned}$$

Exercises

A

For Exercises 1-3, calculate the arc length of $\mathbf{f}(t)$ over the given interval.

1. $\mathbf{f}(t) = (3\cos 2t, 3\sin 2t, 3t)$ on $[0, \pi/2]$

2. $\mathbf{f}(t) = ((t^2+1)\cos t, (t^2+1)\sin t, 2\sqrt{2}t)$ on $[0,1]$

3. $\mathbf{f}(t) = (2\cos 3t, 2\sin 3t, 2t^{3/2})$ on $[0,1]$

4. Parametrize the curve from Exercise 1 by arc length.

5. Parametrize the curve from Exercise 3 by arc length.

B

6. Let $\mathbf{f}(t)$ be a differentiable curve such that $\mathbf{f}(t) \neq \mathbf{0}$ for all t. Show that

$$\frac{d}{dt}\left(\frac{\mathbf{f}(t)}{\|\mathbf{f}(t)\|}\right) = \frac{\mathbf{f}(t) \times (\mathbf{f}'(t) \times \mathbf{f}(t))}{\|\mathbf{f}(t)\|^3}.$$

Exercises 7-9 develop the moving frame field $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ at a point on a curve.

7. Let $\mathbf{f}(t)$ be a smooth curve such that $\mathbf{f}'(t) \neq \mathbf{0}$ for all t. Then we can define the *unit tangent vector* $\mathbf{T}$ by

$$\mathbf{T}(t) = \frac{\mathbf{f}'(t)}{\|\mathbf{f}'(t)\|}.$$

Show that

$$\mathbf{T}'(t) = \frac{\mathbf{f}'(t) \times (\mathbf{f}''(t) \times \mathbf{f}'(t))}{\|\mathbf{f}'(t)\|^3}.$$

8. Continuing Exercise 7, assume that $\mathbf{f}'(t)$ and $\mathbf{f}''(t)$ are not parallel. Then $\mathbf{T}'(t) \neq \mathbf{0}$ so we can define the *unit principal normal vector* $\mathbf{N}$ by

$$\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{\|\mathbf{T}'(t)\|}.$$

Show that

$$\mathbf{N}(t) = \frac{\mathbf{f}'(t) \times (\mathbf{f}''(t) \times \mathbf{f}'(t))}{\|\mathbf{f}'(t)\| \, \|\mathbf{f}''(t) \times \mathbf{f}'(t)\|}.$$

9. Continuing Exercise 8, the *unit binormal vector* B is defined by

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).$$

Show that

$$\mathbf{B}(t) = \frac{\mathbf{f}'(t) \times \mathbf{f}''(t)}{\|\mathbf{f}'(t) \times \mathbf{f}''(t)\|}.$$

Note: The vectors $\mathbf{T}(t)$, $\mathbf{N}(t)$ and $\mathbf{B}(t)$ form a right-handed system of mutually perpendicular unit vectors (called *orthonormal vectors*) at each point on the curve $\mathbf{f}(t)$.

10. Continuing Exercise 9, the *curvature* κ is defined by

$$\kappa(t) = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{f}'(t)\|} = \frac{\|\mathbf{f}'(t) \times (\mathbf{f}''(t) \times \mathbf{f}'(t))\|}{\|\mathbf{f}'(t)\|^4}.$$

Show that

$$\kappa(t) = \frac{\|\mathbf{f}'(t) \times \mathbf{f}''(t)\|}{\|\mathbf{f}'(t)\|^3} \quad \text{and that} \quad \mathbf{T}'(t) = \|\mathbf{f}'(t)\| \, \kappa(t) \mathbf{N}(t).$$

Note: $\kappa(t)$ gives a sense of how "curved" the curve $\mathbf{f}(t)$ is at each point.

11. Find $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$ and κ at each point of the helix $\mathbf{f}(t) = (\cos t, \sin t, t)$.

12. Show that the arc length L of a curve whose spherical coordinates are $\rho = \rho(t)$, $\theta = \theta(t)$ and $\phi = \phi(t)$ for t in an interval $[a, b]$ is

$$L = \int_a^b \sqrt{\rho'(t)^2 + (\rho(t)^2 \sin^2 \phi(t)) \theta'(t)^2 + \rho(t)^2 \phi'(t)^2} \, dt.$$

Chapter Seven: Vectors and Linear equations

Chapter 1

Two and Three-Dimensional Analytic Geometry. Vectors

1.1 Points in three-dimensional space

The study of the geometry of lines and planes in space provides a good introduction to Linear Algebra. Geometry is a visual subject, but it also has an **analytical** aspect which is the study of geometry using algebra: how geometric problems can be expressed and solved algebraically. This geometry is also called **Cartesian Geometry**, after its founder, the 17th Century philosopher and mathematician René Descartes.

From school you are already familiar with the Cartesian plane. There are an $x-$axis and a $y-$axis meeting at right-angles at the origin O. Every point A in the $x-y$ plane is uniquely represented by an ordered pair (a, b) of real numbers, where a is the $x-$coordinate and b is the $y-$coordinate as in Figure 1.1. We write $A = (a, b)$. Here M is the foot of the perpendicular from A to the $x-$axis and likewise N is the the foot of the perpendicular from A to the $y-$axis. Notice that $M = (a, 0)$ and $N = (0, b)$.

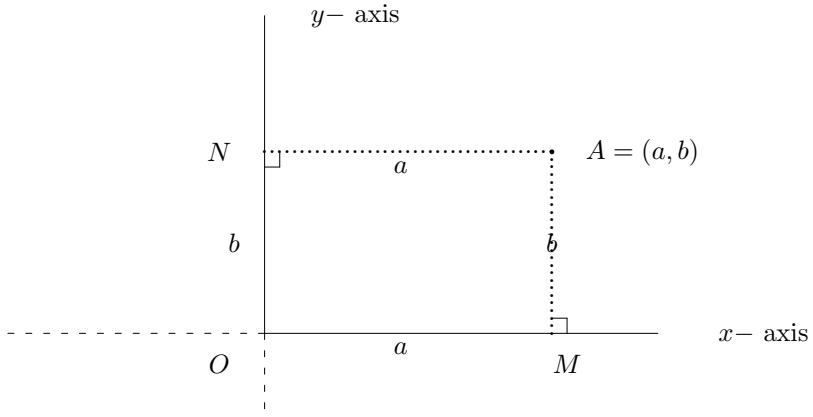


Figure 1.1

In order to represent a point A in space we add the $z-$axis which is perpendicular to both the

$x-$ and $y-$axes as in Figure 1.2. Here $A = (a, b, c)$ is a typical point and a is the $x-$coordinate, b is the $y-$coordinate and c is $z-$coordinate of the point A. In the diagram P is the foot of the perpendicular from A to the $y - z$ plane. Similarly, Q and R are the feet of the perpendiculars from A to the $z - x$ and $x - y$ planes respectively.

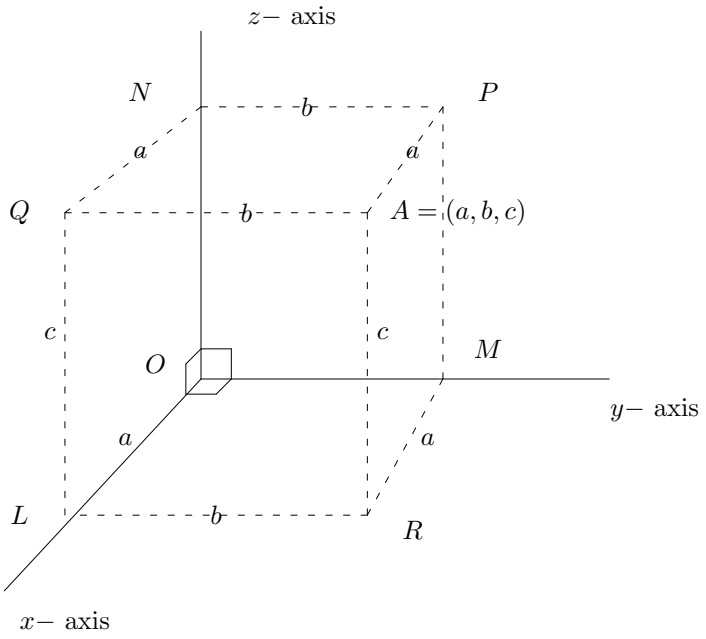


Figure 1.2

1.1.1 The Corkscrew Rule and right-handed systems

You wish to open a bottle of wine using a corkscrew. You first turn the corkscrew so that it enters the cork. As observed by the opener the turning is *clockwise*, but observed from the underside of the bottle the sense is reversed: it is *anticlockwise*. The direction in which a corkscrew moves as it is turned is given by the *Corkscrew Rule*. In Figure 1.2, rotating from the $x-$axis to the $y-$axis and applying this rule produces motion along the (positive) $z-$axis. Rotating from the $y-$axis to the $z-$axis and applying the same rule gives motion along the (positive) $x-$axis. Finally, rotating from the $z-$axis to the $x-$axis and applying the rule produces motion along the (positive) $y-$axis. The axes x, y, z (in that order) are said to form a *right-hand system*. This can be represented by the scheme

$$x \xrightarrow{\quad z \quad} y \xrightarrow{\quad x \quad} z \xrightarrow{\quad y \quad} x$$

Have a look at yourself in the mirror while opening a bottle of wine and observe that what is clockwise for you is anticlockwise for the person in the mirror. Thus a right-hand system of axes observed in a mirror does not follow the corkscrew rule and is called a *left-hand system*. The usual convention is to use only right-hand systems.

1.1.2 Displacements and vectors

Consider two numbers a and b on the real line $\Re$. The difference $c = b - a$ is the **displacement** from a to b. Given any two numbers from a, b and c, the third is uniquely determined by the

equation $c = b - a$. For example, let $a = -2$ and $b = 3$. The displacement from a to b is $c = 3 - (-2) = 5$. The displacement from 6 to 11 is also 5.
Let $A = (a_1, a_2, a_3)$, and $B = (b_1, b_2, b_3)$ be two points in space. The **displacement** from A to B is denoted by $\underline{AB}$ and is given by

$$\underline{AB} = (b_1 - a_1, b_2 - a_2, b_3 - a_3) \tag{1.1}$$

1.1.3 Units and distance

The unit can be any agreed upon distance, like the meter or kilometer. Units (of length, time) are important in applications but for the most part we won't specify them.
The **distance** of $A = (a, b, c)$ from the origin O is given by

$$|\underline{OA}| = \sqrt{a^2 + b^2 + c^2}$$

To see this, see Figure 1.2 and use the theorem of Pythagoras:

$$\begin{aligned} |\underline{OA}|^2 &= |\underline{OR}|^2 + |\underline{RA}|^2 \\ &= |\underline{OL}|^2 + |\underline{LR}|^2 + |\underline{RA}|^2 \\ &= a^2 + b^2 + c^2 \end{aligned}$$

The **distance** from A to B is written as $|\underline{AB}|$ and is given by

$$|\underline{AB}| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2} \tag{1.2}$$

and is also called the **length** or **magnitude** of $\underline{AB}$.

Because the displacement $\underline{AB}$ has both direction and magnitude it is called a **vector** and pictured as an **arrow** going from A to B. The **tail** of the arrow is at A and its **head** is at B as in Figure 1.3.
As an example, let $A = (-1, 2.3, -4.5)$ m and $B = (2.1, 3.6, -2.5)$ m. Then
$\underline{AB} = (2.1 - (-1), 3.6 - 2.3, -2.5 - (-4.5)) = (3.1, 1.3, 2.0)$ m.

Two displacement vectors $\underline{AB}$ and $\underline{CD}$ are equal if they are equal as **displacements**. Hence if $C = (-3.5, 1.1, 3.3)$ and $D = (-.4, 2.4, 5.3)$ and A, B are as above then
$\underline{CD} = (-.4, 2.4, 5.3) - (-3.5, 1.1, 3.3) = (3.1, 1.3, 2.0) = \underline{AB}$.

The magnitude of this vector is $|\underline{AB}| = |\underline{CD}| = \sqrt{(3.1)^2 + (1.3)^2 + (2.0)^2} = 3.9115$ m.
Various vectors (displacement, velocity, acceleration, force) play an important role in mechanics and have their own units. For example, if positions are measured in meters and time is measured in seconds, a velocity vector will have units ms^{-1}. So if a particle moves in such a way that each second it displaces $(-1, 2, 3)$ meters, we say its velocity is constantly $(-1, 2, 3)$ ms^{-1}.

Depending on the geometric interpretation, the tail (head) of displacement vectors may be at any point. Vectors are often written as $\underline{u} = (u_1, u_2, u_3)$, $\underline{v} = (v_1, v_2, v_3)$, $\underline{a} = (a_1, a_2, a_3)$, $\underline{b} = (b_1, b_2, b_3)$ etc.

We now have **three** ways of expressing the position of a point A in space. If A has coordinates $\underline{a} = (a_1, a_2, a_3)$, then

$$A = (a_1, a_2, a_3) = \underline{a} = \underline{OA} \tag{1.3}$$

Although equation (1.3) identifies a point with its position, geometrically speaking we like to distinguish the point itself from its position. The vector $\underline{OA}$ is called the **position vector**

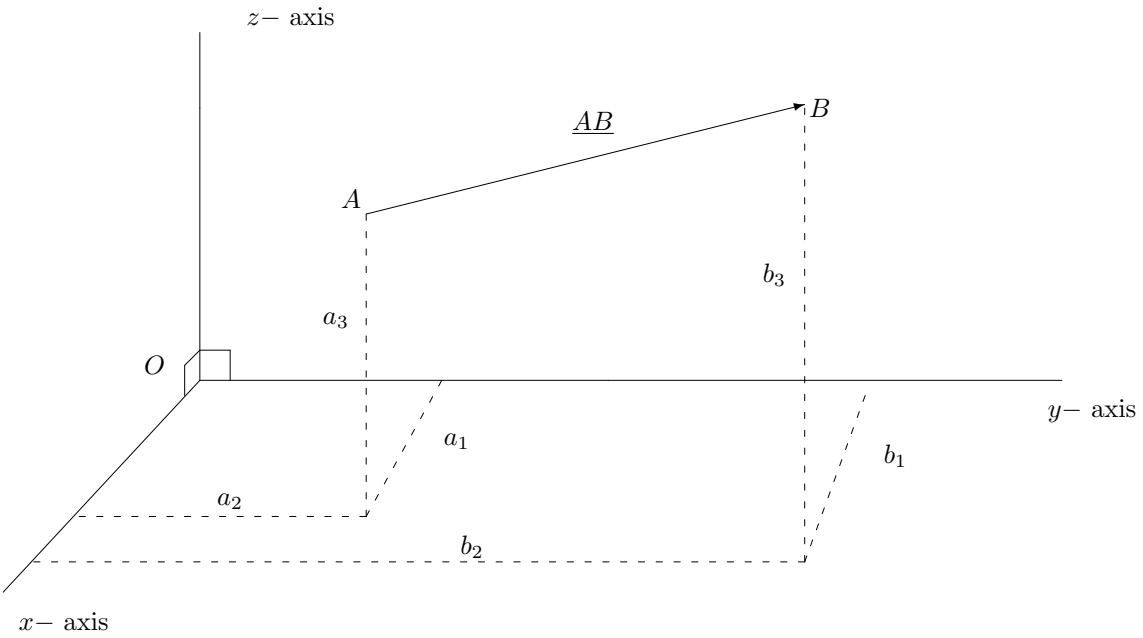


Figure 1.3

of the point A. When we say "$\underline{a}$ is the position vector of point A" it is understood that the tail of $\underline{a}$ is at the origin and its head is at the point A.
Our convention is that we use capital letters $A, B, \ldots$ to denote points in space.
Given the point C, there is a unique point D in space such that $\underline{CD} = \underline{AB}$ (see Exercise 2, No.4). There is also a unique point D such that $\underline{OD} = \underline{AB}$ and we may write $D = \underline{AB}$. In that case the tail of D *can only be* O, and it is better to use the notation $\underline{OD}$ to bring out the vectorial nature of the displacement from O to D.

1.1.4 More systematic notational convention

Instead of speaking of the $x-$, $y-$ and $z-$axis, we also refer to these as the x_1-, x_2- and x_3-axis respectively. Accordingly, given vectors $\underline{a}$, $\underline{b}$, $\underline{c}$,... it will be convenient to assume (unless otherwise stated) that $\underline{a} = (a_1, a_2, a_3)$, $\underline{b} = (b_1, b_2, b_3)$, $\underline{c} = (c_1, c_2, c_3)$, ... Also by convention, points A, P, Q,... have position vectors $\underline{a}$, $\underline{p}$, $\underline{q}$,...., unless otherwise specified.

1.1.5 Addition of vectors

Let $\underline{u} = (u_1, u_2, u_3)$ and $\underline{v} = (v_1, v_2, v_3)$ be two vectors. Their *sum* $\underline{u} + \underline{v}$ is defined by the equation

$$\underline{u} + \underline{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$$

The vector $-\underline{v}$ is defined as $-\underline{v} = (-v_1, -v_2, -v_3)$ and the difference

$$\underline{u} - \underline{v} = \underline{u} + (-\underline{v}) = (u_1 - v_1, u_2 - v_2, u_3 - v_3)$$

The **zero** vector is $\underline{0} = (0, 0, 0)$.

1.1.6 Basic properties of addition of vectors

For all vectors $\underline{u}$, $\underline{v}$ and $\underline{w}$

1. $\underline{u} + (\underline{v} + \underline{w}) = (\underline{u} + \underline{v}) + \underline{w}$ (associative law)
2. $\underline{u} + \underline{v} = \underline{v} + \underline{u}$ (commutative law)
3. $\underline{0} + \underline{u} = \underline{u}$ (the vector $\underline{0} = (0, 0, 0)$ behaves like the number 0)
4. $\underline{u} + (-\underline{u}) = \underline{0}$

An important result for addition is the following:

1.1.7 Geometrical interpretation: The triangle law for addition of vectors

$$\underline{AB} + \underline{BC} = \underline{AC}$$

To see this, let the position vectors of A, B and C be $\underline{a}$, $\underline{b}$ and $\underline{c}$ respectively. Then $\underline{AB} = \underline{b} - \underline{a}$ and $\underline{BC} = \underline{c} - \underline{b}$. Hence by the properties of addition,

$$\begin{aligned} \underline{AB} + \underline{BC} &= (\underline{b} - \underline{a}) + (\underline{c} - \underline{b}) = (\underline{b} - \underline{b}) + (\underline{c} - \underline{a}) \\ &= \underline{0} + \underline{c} - \underline{a} = \underline{c} - \underline{a} = \underline{AC} \end{aligned}$$

Figure 1.4 illustrates this geometrical result in which it is understood that the head B of $\underline{AB}$ is the tail of $\underline{BC}$ etc.
The point D is chosen so that $\underline{AD} = \underline{BC}$. It follows that $\underline{DC} = \underline{AB}$ (why?). We have a geometric interpretation of the above commutative law (known as the **parallelogram law** for addition of vectors):

$$\underline{BC} + \underline{AB} = \underline{AD} + \underline{DC} = \underline{AC} = \underline{AB} + \underline{BC}$$

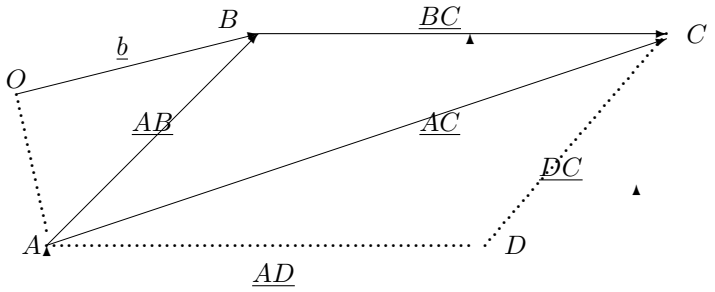


Figure 1.4 $\underline{AC} = \underline{AB} + \underline{BC}$

1.1.8 Multiplication of vectors by scalars. Direction of vectors. Parallel vectors

In contrast to a vector, a **scalar** is just a real number, i.e. an element of the real number system $\Re$. We use α, β, a, b, r, s, t, ... to denote scalars. Note that $\underline{f}$ (f underlined) denotes a vector while f is a scalar.

Let $\underline{u} = (u_1, u_2, u_3)$ be a vector and $\alpha \in \Re$ be a scalar. The product $\alpha\underline{u}$ is defined by

$$\alpha\underline{u} = (\alpha u_1, \alpha u_2, \alpha u_3)$$

The vector $\alpha\underline{u}$ is said to be a (scalar) **multiple** of $\underline{u}$. We usually omit the qualification 'scalar'.

1.1.9 Properties of the product $\alpha\underline{u}$

For all scalars α, β and vectors $\underline{u}$ and $\underline{v}$,

1. $\alpha(\underline{u} + \underline{v}) = \alpha\underline{u} + \alpha\underline{v}$
2. $(\alpha + \beta)\underline{u} = \alpha\underline{u} + \beta\underline{u}$
3. $\alpha(\beta\underline{u}) = (\alpha\beta)\underline{u}$
4. $1\underline{u} = \underline{u}$
5. $\alpha\underline{u} = \underline{0}$ if, and only if, $\alpha = 0$ or $\underline{u} = \underline{0}$
6. $|\alpha\underline{u}| = |\alpha|\,|\underline{u}|$ (the *length* of $\alpha\underline{u}$ is $|\alpha|$ times the length of $\underline{u}$)

The proofs of the first four are left to you. We prove the last:

$$\begin{aligned}
|\alpha\underline{u}| &= |(\alpha u_1, \alpha u_2, \alpha u_3)| \\
&= \sqrt{(\alpha u_1)^2 + (\alpha u_2)^2 + (\alpha u_3)^2} \\
&= \sqrt{\alpha^2 u_1^2 + \alpha^2 u_2^2 + \alpha^2 u_3^2} \\
&= \sqrt{\alpha^2 (u_1^2 + u_2^2 + u_3^2)} \\
&= \sqrt{\alpha^2}\sqrt{u_1^2 + u_2^2 + u_3^2} \\
&= |\alpha| \sqrt{u_1^2 + u_2^2 + u_3^2} \\
&= |\alpha|\,|\underline{u}|
\end{aligned}$$

1.1.10 Geometric interpretation of $\alpha\underline{u}$. Parallel vectors

Since $|\alpha\underline{u}| = |\alpha|\,|\underline{u}|$, multiplication of $\underline{u}$ by the scalar α multiplies the length of $\underline{u}$ by $|\alpha|$.
Let $\underline{u}$ be non-zero, so that it has a definite direction. If $\alpha > 0$ the direction of $\alpha\underline{u}$ is the same as that of $\underline{u}$, while if $\alpha < 0$ the direction of $\alpha\underline{u}$ is opposite to that of $\underline{u}$.

Let $\underline{u}$ and $\underline{v}$ be non-zero vectors. We say they have the **same direction** if $\underline{u} = \alpha\underline{v}$ for some scalar $\alpha > 0$ and **opposite directions** if $\underline{u} = \alpha\underline{v}$ for some $\alpha < 0$. The vector $\underline{u}$ is **parallel to** $\underline{v}$ if $\underline{v}$ is a multiple of $\underline{u}$, that is, $\underline{u} = \alpha\underline{v}$ for some scalar α. Necessarily $\alpha \neq 0$ and we write $\underline{u} \parallel \underline{v}$ to express this relation.
Notice that we only speak of vectors being parallel if they are non-zero.

1.1.11 Properties of parallelism

For all non-zero $\underline{u}$, $\underline{v}$ and $\underline{w}$,

1. $\underline{u} \parallel \underline{u}$
2. $\underline{u} \parallel \underline{v}$ implies $\underline{v} \parallel \underline{u}$
3. $\underline{u} \parallel \underline{v}$ and $\underline{v} \parallel \underline{w}$ together imply $\underline{u} \parallel \underline{w}$.

These properties are geometrically evident, while analytic proofs are left as an exercise. For example, to see property (2) analytically, let $\underline{u} = \alpha\underline{v}$. As $\alpha \neq 0$, $\underline{v} = \frac{1}{\alpha}\underline{u}$ and $\underline{v} \parallel \underline{u}$, as expected.

Example 1.1.1 *Let $\underline{u} = \underline{AB}$, where A is the point $(-1, 7, -4)$ and B is the point $(-4, 11, 1)$, so that $\underline{u} = (-3, 4, 5)$. Let $\underline{v} = \underline{CD}$, where $C = (2, 9, -11)$ and $D = (-7, 21, 4)$, so $\underline{CD} = (-9, 12, 15) = 3\underline{u}$. Hence $\underline{u}$ and $\underline{v}$ have the same direction while $\underline{u}$ and $-\underline{v}$ have opposite directions but are parallel.*

1.1.12 The dot (scalar) product of two vectors

Let $\underline{u} = (u_1, u_2, u_3)$ and $\underline{v} = (v_1, v_2, v_3)$ be any two vectors. Their **dot product** $\underline{u} \cdot \underline{v}$ is defined as

$$\underline{u} \cdot \underline{v} = u_1v_1 + u_2v_2 + u_3v_3 \tag{1.4}$$

Note that the dot product of two vectors is a scalar and not a vector (that is why it is also called the *scalar product*).

1.1.13 Properties of the dot product

For all scalars α, β and vectors $\underline{u}$, $\underline{v}$ and $\underline{w}$,

1. $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$
2. $\alpha\,(\underline{u} \cdot \underline{v}) = (\alpha\underline{u}) \cdot \underline{v} = \underline{u} \cdot (\alpha\underline{v})$
3. $(\alpha\underline{u} + \beta\underline{v}) \cdot \underline{w} = (\alpha\underline{u} \cdot \underline{w}) + (\beta\underline{v} \cdot \underline{w})$
4. $\underline{u} \cdot \underline{u} = |\underline{u}|^2$, so $|\underline{u}| = +\sqrt{\underline{u} \cdot \underline{u}}$
5. $\underline{u} \cdot \underline{u} \geqslant 0$ and $\underline{u} \cdot \underline{u} = 0$ if, and only if $\underline{u} = \underline{0}$.

1.1.14 Geometric interpretation of the dot product

Let $\underline{u} = \underline{AB}$ and $\underline{v} = \underline{AC}$ be *non-zero* vectors. Suppose that $\underline{AB}$ and $\underline{AC}$ make an angle θ between them at the vertex A, where $0 \leq \theta \leq 180$ in degrees, or $0 \leq \theta \leq \pi$ in radians (recall that 180 degrees $= \pi$ radians). Then

$$\underline{u} \cdot \underline{v} = |\underline{u}|\,|\underline{v}| \cos\theta \tag{1.5}$$

To see this result, see Figure 1.5 and use $\underline{AB} + \underline{BC} = \underline{AC}$, so that $\underline{BC} = \underline{AC} - \underline{AB} = \underline{v} - \underline{u}$ and

$$\begin{aligned} |\underline{BC}|^2 &= (\underline{v} - \underline{u}) \cdot (\underline{v} - \underline{u}) \\ &= \underline{v} \cdot \underline{v} + (-\underline{u}) \cdot (-\underline{u}) - \underline{v} \cdot \underline{u} - \underline{u} \cdot \underline{v} \\ &= |\underline{v}|^2 + |\underline{u}|^2 - 2\underline{u} \cdot \underline{v} \end{aligned}$$

Then by the cosine rule (which applies whether θ is acute or obtuse),

$$\begin{aligned} |\underline{BC}|^2 &= |\underline{AB}|^2 + |\underline{AC}|^2 - 2\,|\underline{AB}|\,|\underline{AC}| \cos\theta \\ |\underline{v}|^2 + |\underline{u}|^2 - 2\underline{u} \cdot \underline{v} &= |\underline{u}|^2 + |\underline{v}|^2 - 2\,|\underline{u}|\,|\underline{v}| \cos\theta \quad \text{(from the previous equation)} \end{aligned}$$

Cancelling $|\underline{v}|^2 + |\underline{u}|^2$ from both sides leads to the required result (1.5).

Example 1.1.2 *Let $A = (1, 0, 1)$, $B = (2, 0, 3)$, $C = \left(4, \sqrt{10}, 2\right)$ and $D = (2, 2, -2)$. Find the angle θ between $\underline{AB}$ and $\underline{AC}$. Decide if the angle ϕ between $\underline{AB}$ and $\underline{AD}$ is acute or obtuse.*

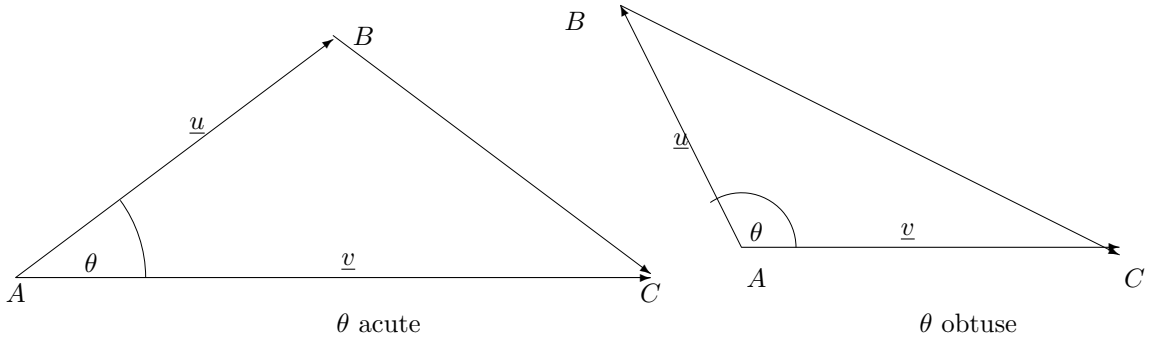


Figure 1.5

Solution: $\underline{AB} = (2,0,3) - (1,0,1) = (1,0,2)$ and $\underline{AC} = \left(4,\sqrt{10},2\right) - (1,0,1) = \left(3,\sqrt{10},1\right)$. Hence

$$\cos\theta = \frac{\underline{AB}\cdot\underline{AC}}{|\underline{AB}|\,|\underline{AC}|} = \frac{(1,0,2)\cdot\left(3,\sqrt{10},1\right)}{\sqrt{1^2+0^2+2^2}\sqrt{3^2+\left(\sqrt{10}\right)^2+1^2}} = \frac{1}{2}$$

and $\theta = 60$ degrees or $\frac{\pi}{3}$ radians.
Similarly, $\underline{AD} = (1,2,-3)$ and

$$\cos\phi = \frac{\underline{AB}\cdot\underline{AD}}{|\underline{AB}|\,|\underline{AD}|} = \frac{(1,0,2)\cdot(1,2,-3)}{\sqrt{1^2+0^2+2^2}\sqrt{1^2+2^2+(-3)^2}} = -\sqrt{\frac{5}{14}}$$

Since the dot product is negative the angle ϕ is obtuse.

1.1.15 Some further properties of the dot product

1. The non-zero vectors $\underline{AB}$ and $\underline{AC}$ are at right angles (are perpendicular $\underline{AB} \perp \underline{AC}$) if, and only if, $(\underline{AB})\cdot(\underline{AC}) = 0$.
   Such vectors are also called mutually **orthogonal**.
2. $|\underline{u}\cdot\underline{v}| \le |\underline{u}|\,|\underline{v}|$ for any vectors $\underline{u}$ and $\underline{v}$ (Cauchy-Schwartz inequality).
3. $|\underline{u}+\underline{v}| \le |\underline{u}| + |\underline{v}|$ for any vectors $\underline{u}$ and $\underline{v}$ (Cauchy's inequality).

The first result follows from the fact that $\underline{AB} \perp \underline{AC}$ if, and only if, $\cos\theta = 0$.
The second result is left as an exercise.
To see the third result, use (2) and consider

$$\begin{aligned}
|\underline{u}+\underline{v}|^2 &= (\underline{u}+\underline{v})\cdot(\underline{u}+\underline{v}) \\
&= \underline{u}\cdot\underline{u} + \underline{v}\cdot\underline{v} + 2\underline{u}\cdot\underline{v} \\
&\le \underline{u}\cdot\underline{u} + \underline{v}\cdot\underline{v} + 2\,|\underline{u}|\,|\underline{v}| \\
&= |\underline{u}|^2 + |\underline{v}|^2 + 2\,|\underline{u}|\,|\underline{v}| \\
&= (|\underline{u}| + |\underline{v}|)^2
\end{aligned}$$

As $|\underline{u}+\underline{v}|^2 \le (|\underline{u}| + |\underline{v}|)^2$, it follows that $|\underline{u}+\underline{v}| \le |\underline{u}| + |\underline{v}|$.

Example 1.1.3 *Find a non-zero vector perpendicular to both* $\underline{u} = (2, 1, -1)$ *and* $\underline{v} = (3, 3, -1)$.

Solution:

Let $\underline{x} = (x_1, x_2, x_3)$ satisfy $\underline{x} \perp \underline{u}$ and $\underline{x} \perp \underline{v}$. Then

$$\begin{aligned} \underline{x} \cdot \underline{u} &= 2x_1 + 1x_2 + (-1)\, x_3 = 0 \\ \underline{x} \cdot \underline{v} &= 3x_1 + 3x_2 + (-1)\, x_3 = 0 \end{aligned}$$

By subtraction $-x_1 - 2x_2 = 0$, or $x_1 = -2x_2$. Thus from the first equation, $x_3 = 2x_1 + x_2 = -3x_2$. We may let x_2 be any non-zero number, e.g. $x_2 = 1$. Then

$$\underline{x} = (-2, 1, -3)$$

is perpendicular to $\underline{u}$ and to $\underline{v}$.
We will see later why it is always possible to find a vector $\underline{x} \neq \underline{0}$ perpendicular to two non-zero vectors. This fact will also come into our study of planes (see 1.3.4).

1.1.16 Unit vectors. Components. Projections

A vector $\underline{u}$ of length 1 is called a **unit vector**. For example, $\underline{u} = \left(\frac{3}{5}, 0, -\frac{4}{5}\right)$ is a unit vector.

A vector $\underline{c} \neq \underline{0}$ defines the vector

$$\widehat{\underline{c}} = \frac{1}{|\underline{c}|}\underline{c}$$

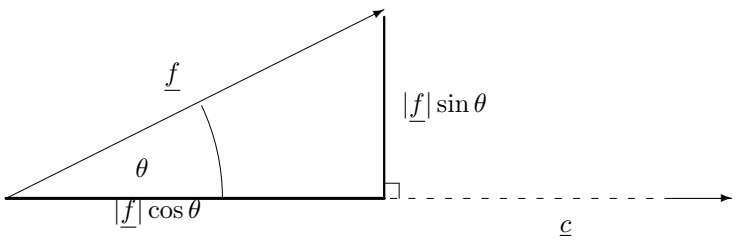
which is the unique **unit vector having the same direction as** $\underline{c}$. To see that $\widehat{\underline{c}}$ has length 1 we note

$$|\widehat{\underline{c}}|^2 = \widehat{\underline{c}} \cdot \widehat{\underline{c}} = \left(\frac{1}{|\underline{c}|}\underline{c}\right) \cdot \left(\frac{1}{|\underline{c}|}\underline{c}\right) = \frac{1}{|\underline{c}|^2}\underline{c} \cdot \underline{c} = \frac{|\underline{c}|^2}{|\underline{c}|^2} = 1$$

The **component** of the vector $\underline{f}$ in the direction of (or along) the non-zero vector $\underline{c}$ is defined as $\underline{f} \cdot \widehat{\underline{c}}$.
Suppose that $\underline{f} \neq \underline{0}$ and that the angle between $\underline{f}$ and $\underline{c}$ is θ. Then

$$\underline{f} \cdot \widehat{\underline{c}} = \left|\underline{f}\right| \cos\theta \tag{1.6}$$

To see equation (1.6), use equation (1.5):



$(\underline{f} \neq \underline{0})$

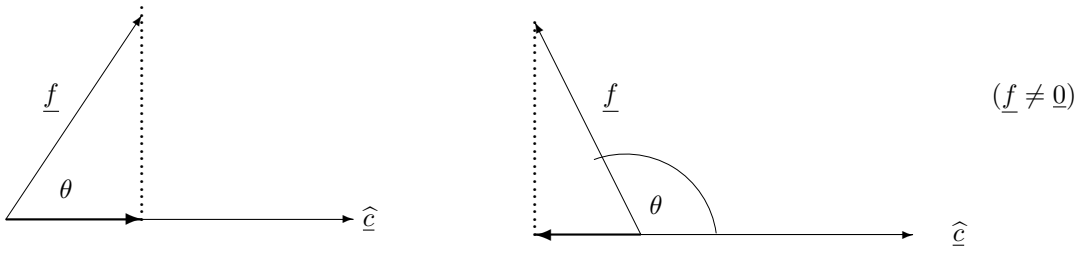
Figure 1.6

$$\underline{f}\cdot\widehat{\underline{c}}=\underline{f}\cdot\frac{1}{|\underline{c}|}\underline{c}=\frac{\underline{f}\cdot\underline{c}}{|\underline{c}|}=\frac{\left|\underline{f}\right||\underline{c}|\cos\theta}{|\underline{c}|}=\left|\underline{f}\right|\cos\theta$$

(See Figure 1.6). Notice that if $\frac{\pi}{2}<\theta\leq\pi$ then the component (1.6) is negative (as is only natural).
The **projection** of $\underline{f}$ on $\underline{c}$ is the vector

$$\left(\underline{f}\cdot\widehat{\underline{c}}\right)\widehat{\underline{c}}=\frac{\underline{f}\cdot\underline{c}}{|\underline{c}|^2}\underline{c} \tag{1.7}$$

Figure 1.7 shows two cases of the projection of $\underline{f}\neq\underline{0}$ on $\underline{c}$: when $\cos\theta>0$ and when $\cos\theta<0$.



Heavy arrow = projection of $\underline{f}$ on $\underline{c}$

Figure 1.7

Let $\underline{d}\parallel\underline{c}$. Then the projection of $\underline{f}$ on $\underline{c}$ is the same as the projection of $\underline{f}$ on $\underline{d}$. See Exercise 2, No.13a.
The situation with components is slightly different: If $\underline{d}$ and $\underline{c}$ have the same direction then the components of $\underline{f}$ along $\underline{c}$ and $\underline{d}$ are the same. If $\underline{d}$ and $\underline{c}$ have opposite directions then the components of $\underline{f}$ along $\underline{c}$ and $\underline{d}$ are the same apart from sign - one is the negative of the other.

We return to projections in subsection 1.2.5.

Example 1.1.4 *1. If we imagine that $\underline{c}$ is pointing due East and $\underline{f}$ is a displacement, then $\left(\underline{f}\cdot\widehat{\underline{c}}\right)\widehat{\underline{c}}$ is the Easterly displacement, while $\left|\underline{f}\right|\cos\theta$ is the Easterly component and $\left|\underline{f}\right|\sin\theta$ is the Northerly component $\underline{f}$.*

*You may think of $\underline{f}$ as a vector representing a **force** that is dragging a box placed at its tail. Then $\left|\underline{f}\right|\sin\theta$ is the lifting effect of $\underline{f}$ while $\left|\underline{f}\right|\cos\theta$ is the horizontal dragging effect of the force.*

2. Let $\underline{f}=(11,-3,4)$ and $\underline{c}=(5,-1,-2)$. The length of $\underline{c}$ is $|\underline{c}|=\sqrt{5^2+(-1)^2+(-2)^2}=\sqrt{30}$.

The unit vector $\widehat{\underline{c}}$ in the direction of $\underline{c}$ is

$$\widehat{\underline{c}}=\frac{1}{\sqrt{30}}(5,-1,-2)$$

The component of $\underline{f}$ along $\underline{c}$ is

$$\underline{f}\cdot\widehat{\underline{c}}=(11,-3,4)\cdot\frac{1}{\sqrt{30}}(5,-1,-2)=\frac{5}{3}\sqrt{30}$$

and the projection of $\underline{f}$ *on* $\underline{c}$ *is*

$$\left(\underline{f}\cdot\widehat{\underline{c}}\right)\widehat{\underline{c}}=\frac{5}{3}\sqrt{30}\left(\frac{1}{\sqrt{30}}\left(5,-1,-2\right)\right)=\frac{5}{3}\left(5,-1,-2\right)$$

1.1.17 The vectors $\underline{i}$, $\underline{j}$, $\underline{k}$

The vectors $\underline{i}$, $\underline{j}$ and $\underline{k}$ are defined as $\underline{i}=(1,0,0)$, $\underline{j}=(0,1,0)$, and $\underline{k}=(0,0,1)$. They have unit length and are directed along the $x-$, $y-$ and $z-$axis respectively. It is clear that they are also mutually orthogonal and that any vector $\underline{a}=(a_1,a_2,a_3)$ has a unique expression

$$\underline{a}=a_1\underline{i}+a_2\underline{j}+a_3\underline{k} \tag{1.8}$$

The scalars a_1, a_2 and a_3 are the components of $\underline{a}$ in the directions $\underline{i}$, $\underline{j}$ and $\underline{k}$ respectively. See also No.15 of Exercise 2 below.

1.1.18 Linear combinations

Let $\underline{a}$ and $\underline{b}$ be two vectors and λ and μ two scalars. The vector $\underline{x}=\lambda\underline{a}+\mu\underline{b}$ is called a **linear combination** of $\underline{a}$ and $\underline{b}$ with coefficients λ and μ. Similarly, for scalars r, s and t, the vector $r\underline{a}+s\underline{b}+t\underline{c}$ is a linear combination of $\underline{a}$, $\underline{b}$, and $\underline{c}$ with coefficients r, s and t.

Equation (1.8) shows that every vector is a linear combination of $\underline{i}$, $\underline{j}$ and $\underline{k}$. If $\underline{a}=(-1,2,3)$, $\underline{b}=(5,-7,1)$, $\underline{c}=(-4,8,-1)$ then

$$\frac{1}{2}\underline{a}+\frac{2}{3}\underline{b}+\left(-\frac{5}{6}\right)\underline{c}=\left(\frac{37}{6},-\frac{31}{3},3\right)$$

is a linear combination of $\underline{a}$, $\underline{b}$, and $\underline{c}$

Exercise 2 *1. In Figure 1.2 express* $\underline{ON}$, $\underline{MP}$, $\underline{OM}$, $\underline{LP}$, $\underline{NR}$, $\underline{NM}$ *and* $\underline{OP}$ *as vectors involving a, b and c. Write down these vectors as linear combinations of* $\underline{i}$, $\underline{j}$ *and* $\underline{k}$ *and find their lengths in terms of a, b and c.*

2. Interpret the equation $(\underline{r}-(2,-1,5))\cdot(\underline{r}-(2,-1,5))=49$ *geometrically.*

3. Let $\underline{u}=(1,3,-2)$, $\underline{v}=(2,0,1)$ *and* $\underline{w}=(3,-3,4)$.

(a) Find the linear combinations $\underline{u}+3\underline{v}-2\underline{w}$, $(-5)\,\underline{u}+2\underline{v}+3\underline{w}$ *and* $\frac{3}{4}\underline{u}-\frac{1}{2}\underline{v}+\frac{1}{4}\underline{w}$.

(b) If the tail of $\underline{u}$ *is the point* $A=(-2,-11,13)$, *what is the head of* $\underline{u}$*? If the head of* $\underline{w}$ *is* $B=(2,-9,10)$, *what is the tail of* $\underline{w}$*?*

(c) Find coefficients α, β *such that* $\underline{w}=\alpha\underline{u}+\beta\underline{v}$.

(d) Is $3\,(3,-1,3)$ *a linear combination of* $\underline{u}$, $\underline{v}$ *and* $\underline{w}$*?*
This means you must look for scalars p, q, r such that $p\underline{u}+q\underline{v}+r\underline{w}=3\,(3,-1,3)$, *and this involves three equations in three unknowns.*

4. (Generalization of No.3b). Let $\underline{u}$ *be a given vector. Show that if the tail A of* $\underline{u}$ *is given then the head B of* $\underline{u}$ *is uniquely determined. State and prove a similar result if the head of* $\underline{u}$ *is given. Draw a picture.*

5. See Figure 1.8. Imagine that in each case the axes are part of a room in which you are standing and looking at the corner indicated. Decide which of the systems of mutually orthogonal axes are left-handed and which right-handed. If you are outside the room, what would your answer be?

6. Let A_0,...,A_9 *be ten points in space and put* $\underline{a}_m=\underline{A_{m-1}A_m}$ *for* $m=1,\cdots,9$ *and* $\underline{a}_{10}=\underline{A_9A_0}$. *Find the sum* $\underline{a}_1+\cdots+\underline{a}_{10}$.

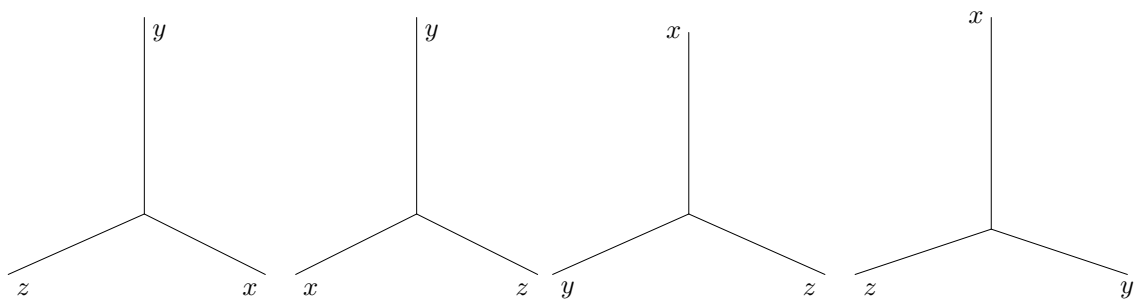


Figure 1.8

7. *Complete the proofs in statements 1.1.6, 1.1.9 and 1.1.11 above.*

   Hint 3 *These properties just reflect similar properties of numbers. For example, the commutative law of addition of vectors is proved by*

$$\begin{aligned} \underline{u}+\underline{v} &= (u_1+v_1, u_2+v_2, u_3+v_3) \\ &= (v_1+u_1, v_2+u_2, v_3+u_3) \\ &= \underline{v}+\underline{u} \end{aligned}$$

   The proof of the first item of 1.1.9 is

$$\begin{aligned} \alpha(\underline{u}+\underline{v}) &= \alpha(u_1+v_1, u_2+v_2, u_3+v_3) \\ &= (\alpha(u_1+v_1), \alpha(u_2+v_2), \alpha(u_3+v_3)) \\ &= (\alpha u_1+\alpha v_1, \alpha u_2+\alpha v_2, \alpha u_3+\alpha v_3) \\ &= \alpha\underline{u}+\alpha\underline{v} \end{aligned}$$

8. *Complete the proofs of subsections 1.1.13 and 1.1.15.*

9. *Deduce that a dot product* $(\alpha\underline{r}+\beta\underline{s})\cdot(\gamma\underline{u}+\delta\underline{v})$ *of linear combinations multiplies out just like in ordinary arithmetic and conclude that* $|\alpha\underline{u}+\beta\underline{v}|^2 = \alpha^2|\underline{u}|^2+\beta^2|\underline{v}|^2+2\alpha\beta\underline{u}\cdot\underline{v}$.

10. *Show that if the vectors* $\underline{u}$ *and* $\underline{v}$ *in equation (1.5) are unit vectors, then* $\cos\theta = \underline{u}\cdot\underline{v}$.

11. *Show that the non-zero vectors* $\underline{u}$ *and* $\underline{v}$ *are parallel if, and only if,* $|\underline{u}\cdot\underline{v}| = |\underline{u}|\,|\underline{v}|$.

    Solution*: The vectors are parallel if, and only if, the angle between them is* 0 *or* π *radians.*

12. *Let* $\underline{a} = (3,-1,2)$ *and* $\underline{b} = (-7,3,-1)$. *Find*

    (a) *The unit vectors* $\widehat{\underline{a}}$ *and* $\widehat{\underline{b}}$.

    (b) *The component of* $\underline{a}$ *along* $\underline{b}$ *and the component of* $\underline{b}$ *along* $\underline{a}$.

    (c) *The projection of* $\underline{a}$ *on* $\underline{b}$ *and the projection of* $\underline{b}$ *on* $\underline{a}$.

13. (a) *If* $\gamma \neq 0$, *show that the projections of* $\underline{f}$ *on* $\underline{c}$ *and* $\underline{f}$ *on* $\gamma\underline{c}$ *are equal.*

    (b) *Show that if* $\underline{f} \neq \underline{0}$, *then the projection of* $\underline{f}$ *on itself is* $\underline{f}$.

14. *Let $\underline{u}$ and $\underline{v}$ be two vectors. The* ***set*** *of all linear combinations of $\underline{u}$ and $\underline{v}$ is called the* ***span*** *of the vectors and is denoted by $sp(\underline{u}, \underline{v})$. (This is discussed more fully in item 2 of subsection 3.1.2 of Chapter 3).*

    Let $\underline{u}_1, \underline{v}_1 \underline{u}_2, \underline{v}_2$ be four vectors. Show that

    $$sp(\underline{u}_1, \underline{v}_1) \subseteq sp(\underline{u}_2, \underline{v}_2)$$

    if, and only if that each of the vectors $\underline{u}_1$ and $\underline{v}_1$ are linear combinations of the vectors $\underline{u}_2$ and $\underline{v}_2$.

15. *Let $\underline{l}$, $\underline{m}$ and $\underline{n}$ be three mutually orthogonal unit vectors. Placing the tails of these vectors at the origin, it is visually clear that we may use these vectors as a frame of reference in place of $\underline{i}$, $\underline{j}$ and $\underline{k}$. (An analytic proof of this will follow from work done in Chapter 3 - See Exercise 78, No.9). Assuming this result, any vector $\underline{a}$ can be expressed as a linear combination*

    $$\underline{a} = \gamma_1 \underline{l} + \gamma_2 \underline{m} + \gamma_3 \underline{n}$$

    Show that necessarily the coefficients γ_1, γ_2, γ_3 are the components (in the above sense) of $\underline{a}$ in the directions $\underline{l}$, $\underline{m}$ and $\underline{n}$ respectively. In fact, $\gamma_1 \underline{l}$, $\gamma_2 \underline{m}$, $\gamma_3 \underline{n}$ are the projections of the vector $\underline{a}$ on $\underline{l}$, $\underline{m}$ and $\underline{n}$ respectively. Conclude that

    $$|\underline{a}|^2 = \gamma_1^2 + \gamma_2^2 + \gamma_3^2$$

    Show that $\underline{l} = \frac{1}{\sqrt{6}}(1, -1, 2)$, $\underline{m} = \frac{1}{\sqrt{3}}(-1, 1, 1)$ and $\underline{n} = \frac{1}{\sqrt{2}}(1, 1, 0)$ is such a system of three mutually orthogonal unit vectors and find the coefficients $\gamma_1, \ldots, \gamma_3$, if $\underline{a} = (-3, 2, 1)$.

    Hint 4 *Take the dot product of both sides of the above equation with each of $\underline{l}$, $\underline{m}$ and $\underline{n}$. The coefficients for the specific example are the components $\underline{a} \cdot \underline{l} = \gamma_1$ etc. A longer procedure to find the coefficients for the specific example is to use the method of 3d above.*

16. *What familiar result from Euclidean Geometry does $|\underline{b} - \underline{a}| \leq |\underline{a}| + |\underline{b}|$ represent?*

17. *Let $\underline{u}$ and $\underline{v}$ have the same length: $|\underline{u}| = |\underline{v}|$. Show that $(\underline{u} - \underline{v}) \cdot (\underline{u} + \underline{v}) = 0$. In other words, if $\underline{u} - \underline{v}$ and $\underline{u} + \underline{v}$ are not zero they are at right angles.*

18. *How are our general results affected (and simplified) when one of the coordinates is restricted to be zero for all vectors under consideration?*

1.2 The straight line

Two distinct points determine a line. Points that lie on one and the same line they are called **collinear**.

Let A, B and C be three distinct points. Then it is geometrically obvious that C lies on the (infinite) line through A and B if, and only if, $\underline{AC}$ is parallel to $\underline{AB}$. We need to show that this axiom is symmetric with regard to A, B and C. In other words, that the statements $\underline{AC} \parallel \underline{AB}$, $\underline{CB} \parallel \underline{CA}$ and $\underline{BA} \parallel \underline{BC}$ are equivalent. It is left as an exercise to prove this equivalence algebraically. See Exercise 5, No.2.
This **axiom** for the collinearity of A, B, and C gives the clue on how to describe algebraically the line ℓ through A and B.

1.2.1 Generic (parametric) representations of ℓ

Let A and B be two distinct points. Then as $t \in \Re$ varies, the point R with position vector

$$\underline{r} = \underline{r}(t) = \underline{OA} + t\underline{AB} \tag{1.9}$$

varies over the entire straight line ℓ through A and B.
As in Figure 1.9, from O jump onto the line at A and then go parallel to $\underline{AB}$ to R. The point $R = A$ is obviously on ℓ and corresponds with $t = 0$. If $R \neq A$ we have $\underline{AR} = t\underline{AB}$ for some $t \neq 0$, so in general

$$\underline{r} = \underline{OR} = \underline{OA} + \underline{AR} = \underline{OA} + t\underline{AB}$$

When $t = 1$ then we are at $R = B$.

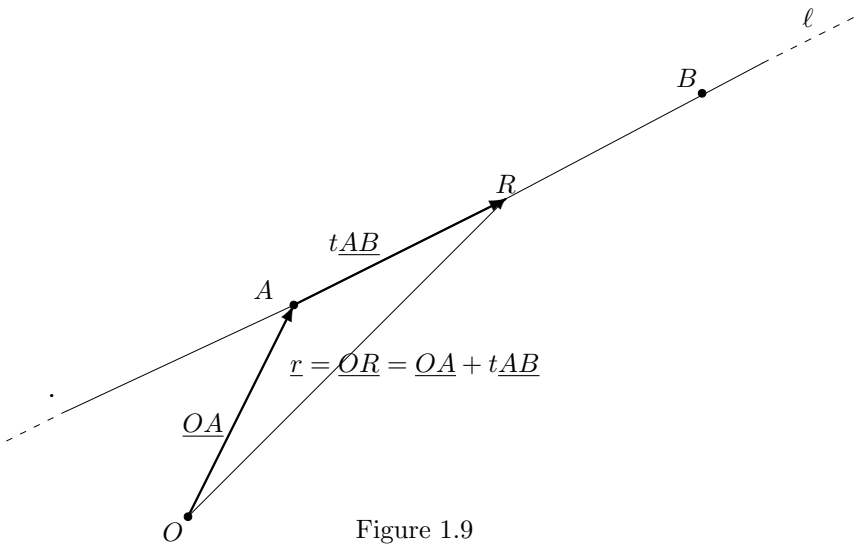


Figure 1.9

As $\underline{r} = \underline{r}(t)$ represents a general point on the line ℓ, the equation (1.9) is called a **generic** or **parametric equation** of the line through A and B. The variable t is called the **parameter**.

1.2.2 Some general observations

In what follows $\underline{a} = \underline{OA}$, $\underline{b} = \underline{OB}$, etc.

1. As t varies over the set $\Re$ of real numbers the point $\underline{r} = \underline{OR} = \underline{a} + t(\underline{b} - \underline{a})$ of equation (1.9) varies continuously over the line ℓ. In this representation of the line ℓ we consider $\underline{b} - \underline{a} = \underline{AB}$ as the **direction** of ℓ. In Figure 1.9 as t increases R moves to the right, as t decreases R moves to the left.

2. We remarked that $t = 0$ corresponds to $R = A$. When $t = 1$ we have $R = B$. If t is between 0 and 1, say $t = \frac{1}{2}$ we get

$$\underline{OR} = \underline{a} + \frac{1}{2}(\underline{b} - \underline{a}) = \frac{1}{2}(\underline{a} + \underline{b})$$

   Thus $\frac{1}{2}(\underline{a} + \underline{b})$ is the position vector of the point **midway** between A and B.

3. More generally, a point

$$\underline{r} = \underline{r}(t) = \underline{a} + t(\underline{b} - \underline{a}) = (1 - t)\underline{a} + t\underline{b}$$

   where $0 \leq t \leq 1$ lies between A and B and the set AB of these points is called the **line segment** joining A and B.
   If $A \neq B$ and $0 < t < 1$, the point $\underline{r}(t)$ is said to lie **strictly between** A and B.
   The line segment AB is a **set** of points and it must be distinguished from the **vector** $\underline{AB}$ and also from the **distance** $|\underline{AB}|$.

4. Let $\underline{u} \neq \underline{0}$ define a direction and suppose $\underline{a}$ is the position vector of the point A. Then a generic equation of the line through A with direction $\underline{u}$ is

$$\underline{r}(t) = \underline{a} + t\underline{u} \tag{1.10}$$

A typical point $\underline{a} + t\underline{u}$ on the line is called a **generic point**.

The **set** ℓ of points $R = \underline{r}(t)$ satisfying a generic equation such as (1.10) as t varies over the real numbers is the **analytic** definition of a straight line.

5. A generic equation is not unique.

   (a) If $C \neq D$ are two points on the line given by equation (1.9), then since these points also determine the line, another generic equation (now with parameter s) is

   $$\underline{r}(s) = \underline{OC} + s\underline{CD}$$

   The parameters t and s are of course related. See Example 1.2.1 of subsection 1.2.3 and No.1b of Exercise 5.

   (b) When do two generic equations like (1.10) define the same line? Let $\underline{u}$ and $\underline{v}$ be non-zero vectors and suppose that $\underline{a} + t\underline{u}$ is a generic equation of line ℓ_1 and $\underline{b} + s\underline{v}$ a generic equation of line ℓ_2.
   Then $\ell_1 = \ell_2$ if, and only if, $\underline{u} \parallel \underline{v}$ and $\underline{a} - \underline{b}$ is a multiple of $\underline{u}$. (See No.5 of Exercise 5).

1.2.3 Some illustrative Examples

Example 1.2.1 *Find two generic equations for the line ℓ passing through $A = (1, 2, 3)$ and $B = (-1, 1, 4)$.*

Solution: One generic equation is

$$\begin{aligned} \underline{r} &= (1,2,3) + t\left((-1,1,4) - (1,2,3)\right) \\ &= (1,2,3) + t(-2,-1,1) = (1-2t, 2-t, 3+t) \end{aligned}$$

Another is

$$\begin{aligned} \underline{r} &= (-1,1,4) + s\left((1,2,3) - (-1,1,4)\right) \\ &= (-1,1,4) + s(2,1,-1) = (-1+2s, 1+s, 4-s) \end{aligned}$$

In the second representation we have used $\underline{r} = \underline{OB} + s\underline{BA}$, with parameter s. The relationship between s and t is $s = 1 - t$, as can be easily seen.

Example 1.2.2 *Find the position vector of the point R lying on the line ℓ of Example 1.2.1 that is a distance $\frac{1}{3}|\underline{AB}|$ from A but on the side of A opposite to that of B.*

Solution: To find R put $t = -\frac{1}{3}$ in $\underline{OR} = \underline{OA} + t\underline{AB}$ to get

$$\begin{aligned} \underline{OR} &= \underline{OA} + \left(-\frac{1}{3}\right)\underline{AB} \\ &= (1,2,3) + \left(-\frac{1}{3}\right)(-2,-1,1) \\ &= \left(\frac{5}{3}, \frac{7}{3}, \frac{8}{3}\right) \end{aligned}$$

Example 1.2.3 *Find the foot Q of the perpendicular from the point $P = (2, -1, 4)$ to the line ℓ of Example 1.2.1. Hence find the shortest distance between P and ℓ.*

Solution:

Let $Q = (1 - 2t, 2 - t, 3 + t)$ be the required foot of the perpendicular from $P = (2, -1, 4)$ to the line ℓ. Then $\underline{PQ} = (-1 - 2t, 3 - t, -1 + t)$ must be perpendicular to the direction $(-2, -1, 1)$ of ℓ:

$$(-1 - 2t, 3 - t, -1 + t) \cdot (-2, -1, 1) = 0$$

This gives $t = \frac{1}{3}$ and so $Q = \underline{q} = \left(\frac{1}{3}, \frac{5}{3}, \frac{10}{3}\right)$. The shortest distance of P from ℓ is thus $\left|\underline{PQ}\right| = \left|\left(-\frac{5}{3}, \frac{8}{3}, -\frac{2}{3}\right)\right| = \frac{1}{3}\sqrt{93}$. (See Figure 1.10).

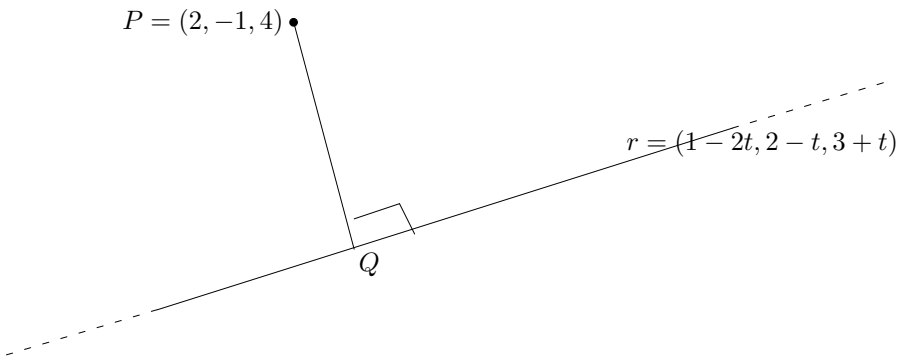


Figure 1.10

Example 1.2.4 *Find the foot Q of the perpendicular from a general point $P = (p_1, p_2, p_3)$ to the line $\underline{r} = \underline{a} + t\underline{u}$ if $\underline{a} = (1, 2, 3)$ and $\underline{u} = (-2, -1, 1)$.*

Solution:

Let t be such that Q has position vector $\underline{a} + t\underline{u}$.

As before, we consider $\underline{PQ} = \underline{a} + t\underline{u} - \underline{p}$ and must solve $\underline{PQ} \cdot \underline{u} = 0$ for t and then substitute this value in $\underline{a} + t\underline{u}$ to obtain Q. We have

$$\left(\underline{a} + t\underline{u} - \underline{p}\right) \cdot \underline{u} = 0 \tag{1.11}$$

Since $\underline{PQ} = (1 - 2t - p_1, 2 - t - p_2, 3 + t - p_3)$, the condition $\underline{PQ} \cdot \underline{u} = 0$ becomes

$$\begin{aligned} & (1 - 2t - p_1, 2 - t - p_2, 3 + t - p_3) \cdot (-2, -1, 1) \\ = \quad & -1 + 6t + 2p_1 + p_2 - p_3 = 0 \end{aligned}$$

This gives $t = \frac{1}{6}(1 - 2p_1 - p_2 + p_3)$. Substituting this value of t in $(1 - 2t, 2 - t, 3 + t)$ gives

$$\underline{q} = \left(\frac{2}{3} + \frac{2}{3}p_1 + \frac{1}{3}p_2 - \frac{1}{3}p_3, \frac{11}{6} + \frac{1}{3}p_1 + \frac{1}{6}p_2 - \frac{1}{6}p_3, \frac{19}{6} - \frac{1}{3}p_1 - \frac{1}{6}p_2 + \frac{1}{6}p_3\right)$$

1.2.4 A general result for the foot of a perpendicular to a line

In Exercise 5, No.11 you are asked to show that the foot Q of the perpendicular from P to the line $\underline{a}+t\underline{u}$ has position vector

$$\underline{q}=\underline{a}+\left(\frac{1}{|\underline{u}|^2}\left(\underline{p}-\underline{a}\right)\cdot\underline{u}\right)\underline{u} \tag{1.12}$$

Example 1.2.5 *Let us apply this formula to the above example 1.2.3 to find the foot of the perpendicular from* $P=(2,-1,4)$ *to the line* $(1,2,3)+t\,(-2,-1,1)$.

Solution:

We have $|\underline{u}|^2=6$ and the foot has position vector

$$\begin{aligned}\underline{q} &= (1,2,3)+\left(\frac{1}{6}\left((2,-1,4)-(1,2,3)\right)\cdot(-2,-1,1)\right)(-2,-1,1)\\ &= \left(\frac{1}{3},\frac{5}{3},\frac{10}{3}\right)\end{aligned}$$

This is the same result as before.

1.2.5 Projections on a line through the origin

When a line ℓ passes through the origin, the foot Q of the perpendicular from P to ℓ is called the **projection** of P on ℓ.
The projection of $\underline{p}$ on the line $t\underline{u}$ is just the projection of $\underline{p}$ on the vector $\underline{u}$. This is equation (1.7) with $\underline{f}=\underline{p}$ and $\underline{c}=\underline{u}$:

$$\underline{q}=\left(\frac{\underline{p}\cdot\underline{u}}{|\underline{u}|^2}\right)\underline{u} \tag{1.13}$$

Example 1.2.6 *Find the projection of* $P=(p_1,p_2,p_3)$ *on the line* $t\,(-2,-1,1)$.

Solution:

In this case $\overrightarrow{PQ}=(-2t-p_1,-t-p_2,t-p_3)$ and we can proceed as before by solving $(-2t-p_1,-t-p_2,t-p_3)\cdot(-2,-1,1)=6t+2p_1+p_2-p_3=0$ for t and substituting into $t\,(-2,-1,1)$. It amounts to the same thing to use (1.13) and the required projection is

$$\begin{aligned}\underline{q} &= \frac{(p_1,p_2,p_3)\cdot(-2,-1,1)}{|(-2,-1,1)|^2}(-2,-1,1)=\frac{-2p_1-p_2+p_3}{6}(-2,-1,1)\\ &= \left(\frac{2}{3}p_1+\frac{1}{3}p_2-\frac{1}{3}p_3,\frac{1}{3}p_1+\frac{1}{6}p_2-\frac{1}{6}p_3,-\frac{1}{3}p_1-\frac{1}{6}p_2+\frac{1}{6}p_3\right)\end{aligned} \tag{1.14}$$

We will return to this example in Chapter 3 (See Example 3.3.1).

Example 1.2.7 *A straight line lies in the* $x-y$ *plane and passes through the origin. The line makes an angle of* 30° *with the positive* $x-$*axis. Find the foot* Q *of the projection from* $P=(p_1,p_2)$ *on the line.*

Solution:

Let $\underline{u}=\left(\cos\frac{\pi}{6},\sin\frac{\pi}{6}\right)=\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)$. The parametric equation of the line is $\underline{r}=t\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)$ and the foot Q of the perpendicular from P to the line is the projection of $\underline{p}=(p_1,p_2)$ on the unit vector $\underline{u}$:

$$\underline{q}=\left(\frac{\sqrt{3}p_1}{2}+\frac{p_2}{2}\right)\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)=\left(\frac{1}{4}\left(3p_1+\sqrt{3}p_2\right),\frac{1}{4}\sqrt{3}p_1+\frac{1}{4}p_2\right) \tag{1.15}$$

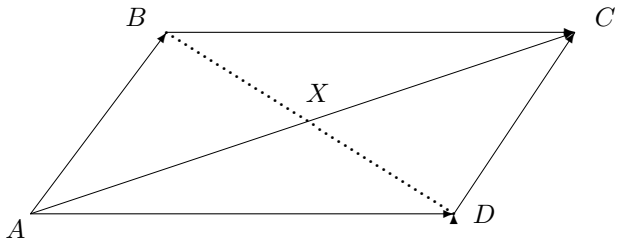
We shall return to this example in subsection 1.4 below and again in Chapter 3 Example 3.2.1.

A well-known result on parallelograms

Example 1.2.8 *Show that the diagonals of a parallelogram bisect each other.*

Solution:

In Figure 1.11 $ABCD$ is a parallelogram with $\underline{AB} = \underline{DC}$ and $\underline{BC} = \underline{AD}$. It is understood that the parallelogram is *non-degenerate* in that no three of A, B, C, D are collinear.



$$\underline{AB} = \underline{DC}$$
$$\underline{AD} = \underline{BC}$$

$$\underline{AX} = \underline{XC}$$

Figure 1.11

Let X be the midpoint of AC so that $\underline{AX} = \underline{XC}$. Then $\underline{AB} + \underline{BX} = \underline{AX} = \underline{XC}$, so $\underline{BX} = \underline{XC} - \underline{AB} = \underline{XC} - \underline{DC} = \underline{XC} + \underline{CD} = \underline{XD}$.

1.2.6 Intersecting, parallel and skew lines

Let ℓ_1 and ℓ_2 given by generic equations $\underline{a}_1 + s\underline{u}_1$ and $\underline{a}_2 + t\underline{u}_2$ respectively be two straight lines. They are parallel if $\underline{u}_1 \parallel \underline{u}_2$.
Exactly one of the following holds (see Exercise 5, No.6* for a rigorous proof):

1. The lines are identical.
2. They are distinct and parallel.
3. They are not parallel and meet in a (unique) point.
4. They are not parallel and do not intersect (they are by definition **skew**).

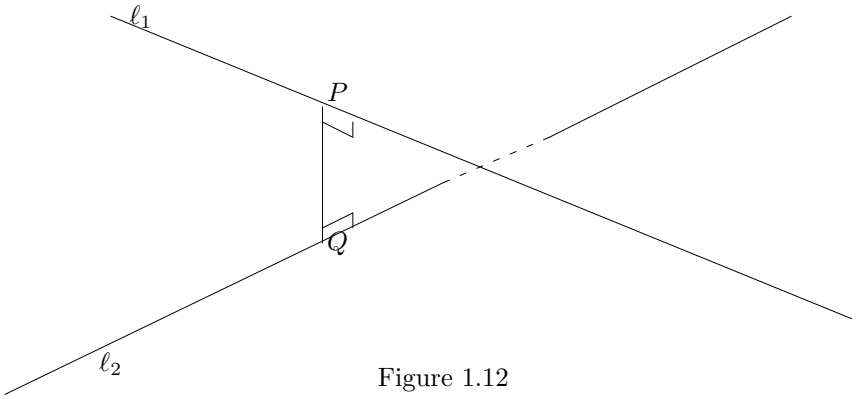


Figure 1.12

A familiar example of skew lines is provided by a road passing under a railway bridge. A train and car can only crash at a level crossing!
In the case of skew lines there will be points P on ℓ_1 and Q on ℓ_2 such that $\underline{PQ}$ is perpendicular to the directions of both lines and the distance $\left|\underline{PQ}\right|$ is minimum. See Figure 1.12.

Example 1.2.9 *Let ℓ_1 be the line through $A = (0,0,0)$ and $B = (1,2,-1)$ and suppose ℓ_2 is the line through $C = (1,1,1)$ and $D = (2,4,3)$. Are the lines skew? In any case, find the shortest distance between them.*

Solution: Generic equations for ℓ_1 and ℓ_2 are $t(1,2,-1)$ and $(1,1,1)+s(1,3,2)$ respectively. Since $(1,2,-1) \nparallel (1,3,2)$, the lines are not parallel (and so cannot possibly be identical). They will meet if for some t and s

$$t(1,2,-1) = (1,1,1) + s(1,3,2)$$

This means that the three equations $t = 1+s$, $2t = 1+3s$ and $-t = 1+2s$ must hold. From the first and last equation, $1+s = -(1+2s)$ and thus $s = -\frac{2}{3}$, $t = \frac{1}{3}$. But then $2t = \frac{2}{3} \neq 1+3\left(-\frac{2}{3}\right)$ and the second equation fails. It follows that ℓ_1 and ℓ_2 are skew.
Let $P = t(1,2,-1)$ and $Q = (1,1,1) + s(1,3,2)$ be the points where PQ is minimum. Then $\underline{PQ} = (1,1,1)+s(1,3,2)-t(1,2,-1)$ is perpendicular to both directions $(1,2,-1)$ and $(1,3,2)$. Hence:

$$\begin{aligned} ((1,1,1) + s(1,3,2) - t(1,2,-1)) \cdot (1,2,-1) &= 2 + 5s - 6t = 0 \\ ((1,1,1) + s(1,3,2) - t(1,2,-1)) \cdot (1,3,2) &= 6 + 14s - 5t = 0 \end{aligned}$$

Solving these two equations gives $s = -\frac{26}{59}$ and $t = -\frac{2}{59}$. Hence,

$$\underline{PQ} = (1,1,1) + s(1,3,2) - t(1,2,-1) = \left(\frac{35}{59}, -\frac{15}{59}, \frac{5}{59}\right)$$

The shortest distance between the lines is $\left|\left(\frac{35}{59}, -\frac{15}{59}, \frac{5}{59}\right)\right| = \frac{5}{\sqrt{59}}$.

Exercise 5 *1. let $A = (1,2,-3)$, $B = (-1,1,2)$, $C = (5,4,-13)$, $D = (-5,-1,12)$ and $E = (-1,-2,3)$. Answer the following questions.*

- *(a) Find a generic equation with parameter t of the straight line ℓ through A and B in the form of equation (1.9).*
- *(b) Is E on ℓ? Show that C and D are on ℓ and find a corresponding generic equation of the line with parameter s. Find the relationship between s and t.*
- *(c) Find two points on ℓ at a distance $\sqrt{24}$ from E.*
- *(d) Find an equation describing the line segment joining A and B. In particular find the midpoint of the segment and also the three points strictly between A and B dividing the segment into four equal parts.*

*2. * Let A, B and C be three distinct points. Show that any of the conditions $\underline{AB} \parallel \underline{AC}$, $\underline{BC} \parallel \underline{AC}$, and $\underline{AB} \parallel \underline{BC}$ are equivalent.*

Hint 6 *Since A, B and C are distinct, $\underline{AB} = \underline{b} - \underline{a}$, $\underline{AC} = \underline{c} - \underline{a}$ and $\underline{BC} = \underline{c} - \underline{b}$ are all non-zero.*
Let p be the statement $\underline{AB} \parallel \underline{AC}$, and q the statement $\underline{BC} \parallel \underline{AC}$ and r the statement $\underline{AB} \parallel \underline{BC}$.

The statement p says there is s with $\underline{b} - \underline{a} = s(\underline{c} - \underline{a})$. From this we get $\underline{c} - \underline{b} = \underline{c} - (\underline{a} + s(\underline{c} - \underline{a})) = \underline{c} - \underline{a} - s(\underline{c} - \underline{a}) = (1-s)(\underline{c} - \underline{a})$. We have just shown $p \Rightarrow q$ (p implies q).
Now substitute C for A, A for B and B for C in $p \Rightarrow q$. You should get $q \Rightarrow r$. Making the same substitutions in $q \Rightarrow r$ you should get $r \Rightarrow p$.

3. *Use the previous criterion to show that any three points satisfying equation (1.10) lie on a straight line (are collinear) and hence that the equation does indeed define a straight line.*

   Hint 7 *Use the results of the previous problem.*

4. *Let the generic equation $\underline{r}(t) = \underline{a} + t\underline{u}$ represent a line ℓ, where $\underline{u} \neq \underline{0}$ and suppose that P with position-vector $\underline{p}$ is a point. Show that P lies on ℓ*

   (a) *if, and only if, $\underline{p} - \underline{a}$ is a multiple of $\underline{u}$.*

   (b) *if, and only if, the equation*

   $$\underline{a} + t\underline{u} = \underline{p}$$

   has a ***unique*** *solution for t.*

   (Compare this with No.14 below for another criterion).

5. ** Let $\underline{u}$ and $\underline{v}$ be non-zero vectors and suppose that $\underline{a} + t\underline{u}$ and $\underline{b} + s\underline{v}$ are generic equations of lines ℓ_1 and ℓ_2 respectively.*
   Show that necessary and sufficient conditions for the generic equations to represent the same set of points (i.e. $\ell_1 = \ell_2$) are

   (i) *$\underline{u} \parallel \underline{v}$ and*

   (ii) *$\underline{a} - \underline{b}$ is a multiple of $\underline{v}$.*

   Using this result show that two lines sharing two distinct points are identical.

6. ** Use the results of the previous exercise to show that the four conditions in subsection 1.2.6 are in fact mutually exclusive and exhaustive.*

7. *Find - but not by simply substituting in equation (1.12) - the point Q of the line ℓ with parametric equation $(1 - 2t, 2 - t, -3 + 5t)$ which is closest to*

   (a) *the origin,*

   Solution. *Solve $(1 - 2t, 2 - t, -3 + 5t) \cdot (-2, -1, 5) = -19 + 30t = 0$. So $t = \frac{19}{30}$ and Q is $(1 - 2t, 2 - t, -3 + 5t) = \left(-\frac{4}{15}, \frac{41}{30}, \frac{1}{6}\right)$*

   (b) *the point $P = (5, 8, 4)$,*

   (c) *the point $P = (-3, 2, 1)$.*

8. *Find the projections of the following points P on the line $(-2t, -t, 5t)$.*

   (a) *$P = (4, 6, 7)$*

   Solution*: We have $((-2t, -t, +5t) - (4, 6, 7)) \cdot (-2, -1, 5) = -21 + 30t = 0$ and $t = \frac{7}{10}$. The projection is $(-2t, -t, +5t) = \left(-\frac{7}{5}, -\frac{7}{10}, \frac{7}{2}\right)$. Alternatively, the projection of P on $(-2, -1, +5)$ is*

   $$\frac{(4, 6, 7) \cdot (-2, -1, +5)}{|(-2, -1, +5)|^2}(-2, -1, +5) = \frac{7}{10}(-2, -1, +5)$$

   (b) *$P = (-1, -2, 3)$.*

   Hint 8 *These can be done in the same way you handled the previous question, but realize that you can use the projection of $\underline{p}$ on the vector $(-2, -1, 5)$.*

9. *Check your results of 7a and 7b using equation (1.12). Can you see any connection between 7a and 8b and also between between 7b and 8a?*

10. *Find the projection of the point P with position vector $\underline{p} = (p_1, p_2, p_3)$ on the line $(-2t, -t, 5t)$.*

11. *Prove the formula (1.12) that finds the foot Q of the perpendicular from $P = (p_1, p_2, p_3)$ onto the line $\underline{a} + t\underline{u}$. Deduce the projection formula (1.13).*

    Hint 9 *Show that the solution to $\left(\underline{a} + t\underline{u} - \underline{p}\right) \cdot \underline{u} = 0$ (equation (1.11) is*

    $$t = -\frac{\left(\underline{a} - \underline{p}\right) \cdot \underline{u}}{\left|\underline{u}\right|^2}$$

12. *Show that the closest the line ℓ with parametric equation $\underline{a} + t\underline{u}$ comes to the point P is*

    $$d = \frac{1}{|\underline{u}|}\sqrt{\left(|\underline{p} - \underline{a}||\underline{u}|\right)^2 - \left(\left(\underline{p} - \underline{a}\right) \cdot \underline{u}\right)^2} \tag{1.16}$$

13. *Show that $|\underline{r}(t)|^2$ in equation (1.10) is a quadratic in t. How does your knowledge of quadratics tie up with the formula (1.12) and equation (1.16)?*

14. *Deduce from equation (1.16) that ℓ passes through the point P if, and only if, $(\underline{p} - \underline{a}) \cdot \underline{u} = \pm|\underline{p} - \underline{a}||\underline{u}|$. Note that geometrically this is clear in view of the formula for the angle between $\underline{p} - \underline{a}$ and $\underline{u}$ if $\underline{p} - \underline{a} \neq \underline{0}$.*

15. ** If t is **time** in seconds and $\underline{u}$ is the displacement in meters per second (the constant **velocity**), then equation (1.10) can be interpreted as the position of (say) an aircraft in meters at time t.*

    (a) *In a situation like this we would be interested in how close the aircraft comes to a certain point P for $t \geq 0$. Find a condition for (1.16) to hold at some time $t \geq 0$. If this condition fails, what is the closest the aircraft gets to P and when does this occur?*

    (b) *How can your findings be used to find the closest distance two aircraft travelling with constant velocities get to each other, assuming that as time $t \geq 0$ increases the distance between the aircraft decreases?*

    Hint 10 *Let $\underline{a} + t\underline{u}$ and $\underline{b} + t\underline{v}$ describe the positions of the two aircraft at time t. Consider the line $\underline{b} - \underline{a} + t\left(\underline{v} - \underline{u}\right)$. You want its closest distance from the origin. By assumption, why is $(\underline{b} - \underline{a}) \cdot (\underline{v} - \underline{u}) < 0$?*

16. *A rhombus is a (non-degenerate) parallelogram all of whose sides have the same length. Show that the diagonals of a rhombus intersect at right angles.*

    Hint 11 *In Example 1.2.8 (Figure 1.11) use $|\underline{AB}| = |\underline{BC}|$ as well as $\underline{AC} = \underline{AB} + \underline{BC}$ and $\underline{AB} + \underline{BD} = \underline{AD}$, so $\underline{BD} = \underline{AD} - \underline{AB} = \underline{BC} - \underline{AB}$. Now use the result of Exercise 2 No.17.*

17. **(Generalization of previous exercise). Let $\underline{a} + t\underline{u}$ and $\underline{b} + s\underline{v}$ represent two straight lines and suppose that $A = \underline{a} + t_1\underline{u}$, $B = \underline{b} + s_1\underline{v}$, $C = \underline{a} + t_2\underline{u}$, and $D = \underline{b} + s_2\underline{v}$ are points on these lines with $A \neq C$ and $B \neq D$.*
    Show that $|\underline{AB}|^2 + |\underline{CD}|^2 = |\underline{BC}|^2 + |\underline{DA}|^2$ if, and only if, the lines are perpendicular.

    Hint 12 *Expand $|\underline{AB}|^2 = (\underline{b} + s_1\underline{v} - (\underline{a} + t_1\underline{u})) \cdot (\underline{b} + s_1\underline{v} - (\underline{a} + t_1\underline{u}))$ etc. and show that $|\underline{AB}|^2 + |\underline{CD}|^2 - |\underline{BC}|^2 - |\underline{DA}|^2 = 2\underline{u} \cdot \underline{v}(t_1 - t_2)(s_2 - s_1)$.*

18. *Let A, B and C be three points not on a line and so forming the distinct vertices of a triangle. The median from A is the line joining A to the midpoint of the opposite side BC. Show that all three medians intersect in a single point M. This point is called the* ***centroid*** *of the triangle and is the* ***centre of mass*** *of three equal masses placed at the vertices of the triangle.*

    Hint 13 *Let $\underline{a}$, $\underline{b}$, and $\underline{c}$, be the position vectors of A, B and C respectively. Then the midpoint of BC has position vector $\frac{1}{2}(\underline{b}+\underline{c})$. A generic equation of the line through A and the midpoint of BC is*

$$\underline{a} + t\left(\frac{1}{2}(\underline{b}+\underline{c}) - \underline{a}\right)$$

    Put $t = \frac{2}{3}$ and simplify the result. You should conclude that the centroid lies two thirds of the way from any vertex along its median toward the opposite side.

19. *(Cartesian equations for a straight line). Let equation (1.10) define a straight line, where $\underline{u} = (u_1, u_2, u_3)$ and each $u_i \neq 0$. Show that $\underline{r} = (x, y, z)$ is on the line if, and only if,*

$$\frac{x-a_1}{u_1} = \frac{y-a_2}{u_2} = \frac{z-a_3}{u_3}$$

    What happens if one or two of the $u_i = 0$? Find Cartesian equations for the line of Question 1a above. Find generic and Cartesian equations for the $x-$, $y-$ and $z-$axis.

20. *Let ℓ_1 be the line through $A = (-1, -2, 6)$ and $B = (3, 2, -2)$ and ℓ_2 the line through $C = (1, 3, 1)$ and $D = (1, 1, 5)$. Solve the following problems.*

    (a) *Show that ℓ_1 and ℓ_2 are skew.*

    (b) *Find points P on ℓ_1 and Q on ℓ_2 such that $\underline{PQ}$ is perpendicular to the directions of both lines. Hence find the shortest distance between ℓ_1 and ℓ_2.*

21. *Let ℓ_1 be the line through $A = (2, -1, 4)$ and $B = (3, 0, 5)$ and ℓ_2 the line through $C = (-1, 0, 1)$ and $D = (-2, 1, 0)$. Show that ℓ_1 and ℓ_2 meet and find the point of intersection.*

22. *How are our general results affected (and simplified) when one of the coordinates (say the $z-$ coordinate) is restricted to be zero? In particular what does equation (1.12) reduce to?*

1.3 Planes

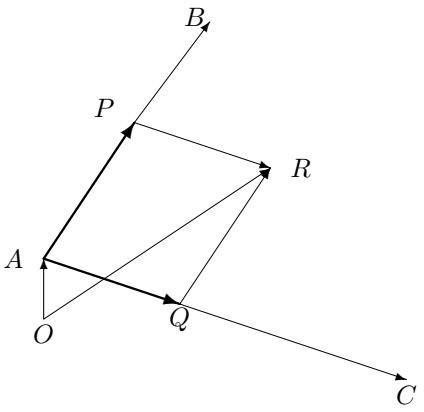
1.3.1 Generic Equation for a plane

Our first object is to find a generic equation for the plane passing through three **non-collinear points** $A = (a_1, a_2, a_3)$, $B = (b_1, b_2, b_3)$ and $C = (c_1, c_2, c_3)$.

If four points lie on a plane, they are called **coplanar**. So if R is a point, our first question is "when are A, B, C and R coplanar?" The answer is suggested by Figure 1.13:
Suppose that R lies in the same plane as A, B and C. Let the line through R parallel to $\underline{AC}$ meet the line through A and B at the point P. Then $\underline{AP} = s\underline{AB}$ for some scalar s. Similarly, let the line through R parallel to $\underline{AB}$ meet the line through A and C at the point Q, so that $\underline{AQ} = t\underline{AC}$ for some scalar t. The figure $APRQ$ forms a parallelogram and

$$\begin{aligned} \underline{OR} &= \underline{OA} + \underline{AR} = \underline{OA} + (\underline{AP} + \underline{PR}) = \underline{OA} + \left(\underline{AP} + \underline{AQ}\right) \\ &= \underline{OA} + s\underline{AB} + t\underline{AC} \end{aligned} \tag{1.17}$$

This equation describes the position vector $\underline{r} = \underline{OR}$ of a general point on the plane through A, B and C and is called **a generic equation** of the plane through A, B and C.



$$s\underline{AB} = \underline{AP} = \underline{QR}$$

$$t\underline{AC} = \underline{AQ} = \underline{PR}$$

Figure 1.13

Example 1.3.1 *Let $A = (1, -1, 2)$, $B = (1, 2, 3)$ and $C = (-1, 0, 1)$ be three given points. They are not collinear since the vectors $\underline{AB} = (0, 3, 1)$ and $\underline{AC} = (-2, 1, -1)$ are not parallel. Hence a generic equation of the plane through A, B and C is*

$$\underline{r} = (1, -1, 2) + s(0, 3, 1) + t(-2, 1, -1)$$

In equation (1.17) the vectors $\underline{u} = \underline{AB}$ and $\underline{v} = \underline{AC}$ are non-zero and non-parallel and are thought of as parallel to the plane described. This suggests the following definition:

1.3.2 Analytic definition of a plane

Let $\underline{u}$ and $\underline{v}$ be two non-zero, non-parallel vectors and $A = \underline{a}$ a given point. Then the **set** Π of points $R = \underline{r}$ satisfying the generic (parametric) equation

$$\underline{r} = \underline{r}(s, t) = \underline{a} + s\underline{u} + t\underline{v} \tag{1.18}$$

describes a plane. The parameters are s and t and we call $\underline{a} + s\underline{u} + t\underline{v}$ a **generic point** of the plane.

1.3.3 Equality of planes

As with lines, a generic equation for a plane is not unique. Let

$$\underline{a}_1 + s_1\underline{u}_1 + t_1\underline{v}_1 \qquad \text{and} \quad \underline{a}_2 + s_2\underline{u}_2 + t_2\underline{v}_2$$

Represent planes Π_1 and Π_2 respectively. Then $\Pi_1 = \Pi_2$ if, and only if,

(i)
$$sp\,(\underline{u}_1, \underline{v}_1) = sp\,(\underline{u}_2, \underline{v}_2)$$

and

(ii) $\underline{a}_1 - \underline{a}_2$ is in this common span.
For the meaning of 'span' see No.14 of Exercise 2 and compare No.5 of Exercise 5.

The proof of the above statement is left as an exercise. See No.11 of Exercise 16 below.

1.3.4 Cartesian Equation of a plane

A vector $\underline{n} = (n_1, n_2, n_3) \neq \underline{0}$ is called a **normal** of the plane defined by equation (1.18) if it is perpendicular to every line lying in the plane. So for a variable point R on the plane we must have $\underline{n} \cdot \underline{AR} = 0$. Thus as in Figure 1.14, all points on the plane with position vector $\underline{r} = (x_1, x_2, x_3)$ must satisfy

$$\underline{n} \cdot (\underline{r} - \underline{a}) = 0$$

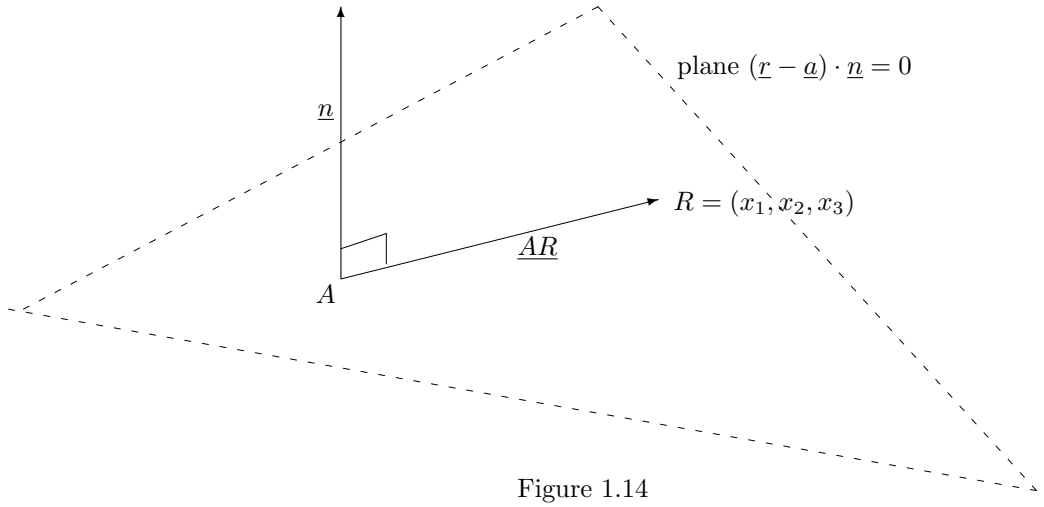


Figure 1.14

This is a **Cartesian equation** of the plane.

More fully this reads

$$\begin{aligned} & n_1x_1 + n_2x_2 + n_3x_3 - (n_1a_1 + n_2a_2 + n_3a_3) \\ = \quad & n_1x_1 + n_2x_2 + n_3x_3 + c = 0 \end{aligned} \tag{1.19}$$

where $c = -(n_1a_1 + n_2a_2 + n_3a_3)$

A Cartesian equation is unique up to non-zero constant multiples of (n_1, n_2, n_3, c) since evidently multiplying n_1, n_2, n_3 and c by $\beta \neq 0$ cannot change a solution (x_1, x_2, x_3).

1.3.5 Finding a normal $\underline{n}$

If $\underline{n}$ is a normal to the plane given by equation (1.18), then in particular $\underline{n} \perp \underline{u}$ and $\underline{n} \perp \underline{v}$. In other words, simultaneously

$$\begin{aligned} n_1u_1 + n_2u_2 + n_3u_3 &= 0 \\ n_1v_1 + n_2v_2 + n_3v_3 &= 0 \end{aligned} \tag{1.20}$$

Conversely, if equation (1.20) holds then $\underline{n}$ must be a normal (see Exercise 16, No.4 below). Because equation (1.20) involves *two* equations for *three* unknowns, a solution $\underline{n} \neq \underline{0}$ can always be found. See Chapter 2, 2.2.15 for a proof of this.

Example 1.3.2 *Find a normal $\underline{n} \neq \underline{0}$ to the plane of Example 1.3.1 and hence its Cartesian equation.*

Solution:

We solve

$$\begin{aligned}(n_1, n_2, n_3) \cdot (0, 3, 1) &= 3n_2 + n_3 = 0 \\ (n_1, n_2, n_3) \cdot (-2, 1, -1) &= -2n_1 + n_2 - n_3 = 0\end{aligned}$$

simultaneously for n_1, n_2 and n_3. (We did much the same thing in Example 1.1.3). From the first equation, $n_3 = -3n_2$. Substituting this into the second equation, we get $n_1 = \frac{1}{2}(n_2 - n_3) = 2n_2$. Thus, provided $n_2 \neq 0$, the vector

$$\underline{n} = (2n_2, n_2, -3n_2) = n_2(2, 1, -3)$$

is perpendicular (normal) to the plane. We may take $n_2 = 1$ and $\underline{n} = (2, 1, -3)$. With $A = (1, -1, 2)$ we get a Cartesian equation for the plane:

$$(2, 1, -3) \cdot ((x_1, x_2, x_3) - (1, -1, 2)) = 2x_1 + x_2 - 3x_3 + 5 = 0$$

1.3.6 Some remarks on planes

Remark 14 *The following statements are geometrically obvious, but analytic proofs can be given after we have studied the subject more deeply, in particular after we ave done Chapter 3. See No.10 in Exercise 78.*

1. Provided $\underline{n} = (n_1, n_2, n_3) \neq \underline{0}$, an equation $n_1x_1 + n_2x_2 + n_3x_3 + c = 0$ always represents a plane with normal $\underline{n}$. (If $\underline{n} = \underline{0}$ the equation represents nothing if $c \neq 0$ and all of space if $c = 0$).
2. By definition, two distinct planes are **parallel** if their normals are parallel. Such planes are a constant positive distance apart and have no points in common.
3. Two non-parallel planes meet in a line.
4. By definition, a **line is parallel to a plane** if it is perpendicular to its normal (draw a picture to see that this is correct). A line not parallel to a plane meets the plane in exactly one point.
5. A line and a point not on it determine a unique plane.
6. There are an infinity of planes containing a given line.

Example 1.3.3 *Find a generic equation for the line of intersection of the planes*

$$\begin{aligned}x_1 + x_2 + x_3 + 1 &= 0 \\ 2x_1 + x_2 - x_3 + 2 &= 0\end{aligned}$$

Solution:

Note that $(1, 1, 1) \nparallel (2, 1, -1)$ so the planes are not parallel. Eliminating x_1 gives $x_2 + 3x_3 = 0$ or $x_2 = -3x_3$. Substituting $x_2 = -3x_3$ into the first (or second) equation gives $x_1 = 2x_3 - 1$. Hence, with $t = x_3$ the parameter,

$$\underline{r} = (x_1, x_2, x_3) = (2t - 1, -3t, t)$$

is the line of intersection of the two planes.

Example 1.3.4 *Find a generic equation of the plane with Cartesian equation*

$$x_1 - 2x_2 + 5x_3 = 10 \tag{1.21}$$

Solution:

By far the easiest solution is

$$\underline{r} = (10 + 2x_2 - 5x_3, x_2, x_3)$$

Here the independent parameters are x_2 and x_3 and they can vary freely over $\Re$. We could of course let $s = x_2$ and $t = x_3$, but this is only an artificial device and changes nothing.

A more long-winded approach is to find three non-collinear points A, B and C on the plane and then proceed as in Example (1.3.1).

Letting $x_2 = x_3 = 0$ gives $x_1 = 10$ and $A = (10, 0, 0)$ as one point. Putting $x_1 = x_2 = 0$ gives $x_3 = 2$ and $B = (0, 0, 2)$. With $x_1 = x_3 = 0$ we have $x_2 = -5$ and $C = (0, -5, 0)$. The points A, B and C do not lie on a line and another generic equation of the plane is

$$\begin{aligned} \underline{r} &= \underline{OA} + s\underline{AB} + t\underline{AC} \\ &= (10 - 10s - 10t, -5t, 2s) \end{aligned}$$

Remark 15 *Instead of A, B and C we could have chosen any three non-collinear points satisfying equation (1.21). The parametric equation will be different but will represent the same plane.*

Example 1.3.5 *Find two non-parallel vectors $\underline{u}$ and $\underline{v}$ both perpendicular to $\underline{n} = (1, -2, 5)$.*

Solution:

Two such vectors are $\underline{u} = \underline{AB}$ and $\underline{v} = \underline{AC}$

Example 1.3.6 *Find the point of intersection of the line through $(1, -1, 0)$ and $(0, 1, 2)$ and the plane*

$$3x_1 + 2x_2 + x_3 + 1 = 0$$

Solution:

For some t the point

$$\begin{aligned} \underline{r} &= (1, -1, 0) + t\left((0, 1, 2) - (1, -1, 0)\right) \\ &= (1, -1, 0) + t\left(-1, 2, 2\right) = (1 - t, -1 + 2t, 2t) \end{aligned}$$

must be on the given plane, i.e.

$$3\left(1 - t\right) + 2\left(-1 + 2t\right) + \left(2t\right) + 1 = 0$$

Thus $t = -\frac{2}{3}$ and the point of intersection is $\left(\frac{5}{3}, -\frac{7}{3}, -\frac{4}{3}\right)$.

1.3.7 Perpendicular from a Point to a Plane. Projections

If Π is a plane and P a point, the problem in this section is to find the foot Q of the perpendicular from P to the plane. See Figure 1.15.
If the plane passes through the origin, the foot Q is called the **projection** of P on the plane.

Example 1.3.7 *Find the foot Q of the perpendicular from the point $P = (4, -1, 3)$ to the plane $x + 2y + z = -4$. Hence find the shortest distance from P to the given plane.*

Solution:

The vector $(1, 2, 1)$ is normal to the plane so the line ℓ defined by $(4, -1, 3) + t\left(1, 2, 1\right) = (4 + t, -1 + 2t, 3 + t)$ passes through P and has direction normal to the plane. Geometry tells

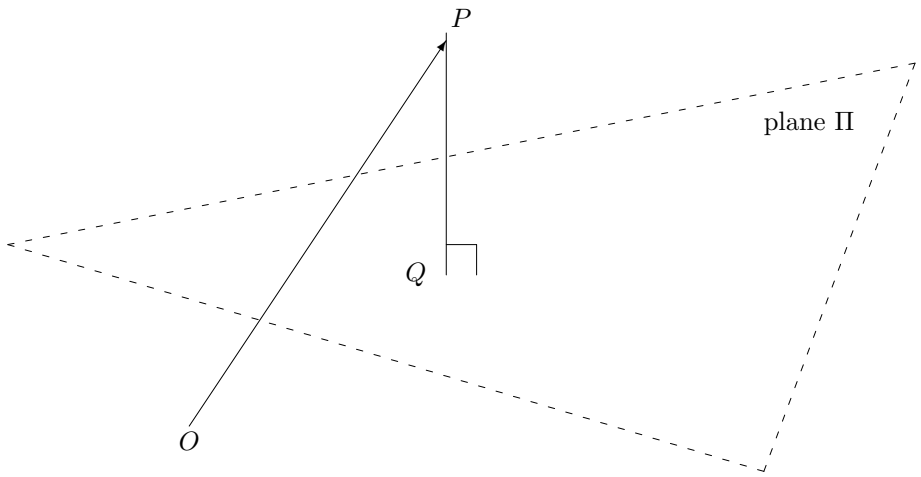


Figure 1.15 Perpendicular Q from P to plane Π

us that the line ℓ hits the plane at the required foot Q of the perpendicular. Then $\left|\underline{PQ}\right|$ will be the shortest distance of P from the plane.

To find t, substitute $(4+t, -1+2t, 3+t)$ into the Cartesian equation of the plane

$$(4+t) + 2(-1+2t) + (3+t) = -4$$

and solve for t. This gives $t = -\frac{3}{2}$ and the position vector of the point Q therefore has position vector

$$\underline{q} = \left(4 + \left(-\frac{3}{2}\right), -1 + 2\left(-\frac{3}{2}\right), 3 + -\frac{3}{2}\right) = \left(\frac{5}{2}, -4, \frac{3}{2}\right)$$

The required shortest distance of P from the plane is

$$\begin{aligned} \left|\underline{PQ}\right| &= \left|\left(\frac{5}{2}, -4, \frac{3}{2}\right) - (4, -1, 3)\right| \\ &= \left|\left(-\frac{3}{2}, -3, -\frac{3}{2}\right)\right| = \sqrt{\frac{27}{2}} \end{aligned}$$

Example 1.3.8 *Find the foot Q of the perpendicular from a general point $P = (p_1, p_2, p_3)$ to*

1. *the above plane $x + 2y + z = -4$, and*
2. *the parallel plane $x + 2y + z = 0$ which passes through the origin.*

Solution:

1. Consider the line $(p_1, p_2, p_3) + t(1, 2, 1) = (p_1 + t, p_2 + 2t, p_3 + t)$. As above, we find where this line hits the given plane. The equation determining this point Q is

$$(p_1 + t) + 2(p_2 + 2t) + (p_3 + t) = -4$$

This gives $t = -\frac{4+p_1+2p_2+p_3}{6}$. Hence Q is

$$\underline{q} = (p_1, p_2, p_3) - \frac{4 + p_1 + 2p_2 + p_3}{6}(1, 2, 1) \tag{1.22}$$

2. We must substitute $(p_1, p_2, p_3) + t(1, 2, 1) = (p_1 + t, p_2 + 2t, p_3 + t)$ into the equation $x + 2y + z = 0$ of the plane:

$$(p_1 + t) + 2(p_2 + 2t) + (p_3 + t) = 0$$

This gives $t = -\frac{p_1+2p_2+p_3}{6}$ and hence

$$\underline{q} = (p_1, p_2, p_3) - \frac{p_1 + 2p_2 + p_3}{6}(1, 2, 1) \tag{1.23}$$

Because the plane $x + 2y + z = 0$ passes through the origin the point Q given by equation (1.23) is the projection of P on the plane.

See Exercise 16, (5), where the theory is taken further.

Exercise 16 *1. Show that $A = (1, 2, 1)$, $B = (0, 1, 3)$ and $C = (-1, 2, 5)$ are not collinear and so define a unique plane Π. Find:*

(a) A generic equation for Π

(b) A normal to Π and hence a Cartesian equation for Π

(c) The foot Q of the perpendicular from $P = (1, 2, 3)$ to Π and so the shortest distance from P to the plane Π

(d) The foot Q of the perpendicular from $P = (p_1, p_2, p_3)$ to Π (see also No.5 below).

(e) The intersection of the line ℓ with parametric form $(1 + t, 3 + 2t, 4 - t)$ and Π, as well as the acute angle between the line and the normal to Π.

(f) A plane containing the point $(4, 5, -3)$ and the line $(1 + t, 3 + 2t, 4 - t)$.

2. Find a generic equation for the plane $2x + 3y - 4z + 5 = 0$.

3. Find generic and Cartesian equations for the $x - y$ and $y - z$ planes.

4. Prove that if $\underline{n}$ satisfies equation (1.20) then it is normal to the plane given by the generic equation (1.18).

5. Follow the method of Examples (1.3.7) and (1.3.8) above to show that the foot Q of the perpendicular from the point P to the plane Π with Cartesian equation $\underline{n} \cdot \underline{r} + c = 0$ is given by

$$\underline{q} = \underline{p} - \left(\frac{c + \underline{n} \cdot \underline{p}}{|\underline{n}|^2}\right)\underline{n} \tag{1.24}$$

Here $\underline{p}$ is the position vector of P.
*Consider also the case $c = 0$. Then Q is called the **projection** of P on the plane*

Hint 17 *Solve*

$$\underline{n} \cdot (\underline{p} + t\underline{n}) + c = 0$$

for t.

6. Use equation (1.24) to show that the shortest distance of the point P from the plane $\underline{n} \cdot \underline{r} + c = 0$ is

$$\frac{|\underline{n} \cdot \underline{p} + c|}{|\underline{n}|}$$

7. Write the equation of a plane $\underline{n} \cdot \underline{p} + c = 0$ in the form $\underline{m} \cdot \underline{p} + d = 0$ where $\underline{m}$ is unit vector.

8. *Let* $\underline{n} \cdot \underline{r} + c_1 = 0$ *and* $\underline{n} \cdot \underline{r} + c_2 = 0$ *represent planes* Π_1 *and* Π_2 *respectively. Show algebraically that* Π_1 *and* Π_2 *are either identical or have no point in common. In the latter case write down a formula for the shortest distance between* Π_1 *and* Π_2 *if* $P = (p_1, p_2, p_3)$ *is a given point on the first plane.*

9. *Find a generic equation of the line of intersection of the planes* $2x + 3y + z + 1 = 0$ *and* $x - 4y + 5z - 2 = 0$.

10. *(Hirst, p.36). In 10a - 10c find the intersection of the three given planes. In each case give a brief geometric description.*

(a)

$$\begin{aligned} 3x - 2y + 2z &= 5 \\ 4x + y - z &= 3 \\ x + y + z &= 6 \end{aligned}$$

(b)

$$\begin{aligned} x + 2y - z &= 3 \\ 2x + 4y - 2z &= 5 \\ x - y + z &= 1 \end{aligned}$$

(c)

$$\begin{aligned} x + 2y - z &= 2 \\ x - y + z &= 1 \\ 3x + 3y - z &= 5 \end{aligned}$$

11. ** Prove the statement of subsection 1.3.3.*

Hint 18 *Imitate the procedure of No.5 of Exercise 5.*

12. *(See Figure 1.16) Let* $\underline{n} \cdot \underline{r} + c = 0$ *and* $\underline{m} \cdot \underline{r} + d = 0$ *represent non-parallel planes* Π_1 *and* Π_2 *respectively. Show that the Cartesian equation of a plane* Π *that contains the line of intersection of* Π_1 *and* Π_2 *has the form*

$$\begin{aligned} & \lambda\left(\underline{n} \cdot \underline{r} + c\right) + \mu\left(\underline{m} \cdot \underline{r} + d\right) \qquad (1.25) \\ = \quad & (\lambda\underline{n} + \mu\underline{m}) \cdot \underline{r} + \lambda c + \mu d = 0 \end{aligned}$$

where λ *and* μ *are two scalars not both of which are zero.*

Hint 19 *First note that because* $\underline{n} \neq \underline{0}$ *and* $\underline{m} \neq \underline{0}$ *are not parallel, the linear combination* $\lambda\underline{n} + \mu\underline{m}$ ***is also non-zero if one or both of*** λ, μ ***are non-zero.*** *In that case equation (1.25) represents a plane* Π. *Note that if (e.g.)* $\lambda \neq 0$ *and* $\mu = 0$ *then* Π. *is just* Π_1. *Now check that a point on* Π_1 *and* Π_2 *is also on* Π.

13. *Find the form of a Cartesian equation of the plane* Π *that contains the intersection of the planes*

$$2x_1 + 3x_2 + x_3 + 1 = 0 \quad \textit{and} \quad x_1 - 4x_2 + 5x_3 - 2 = 0$$

If the plane Π *also contains the point* $(1, -1, 2)$, *find its Cartesian equation.*

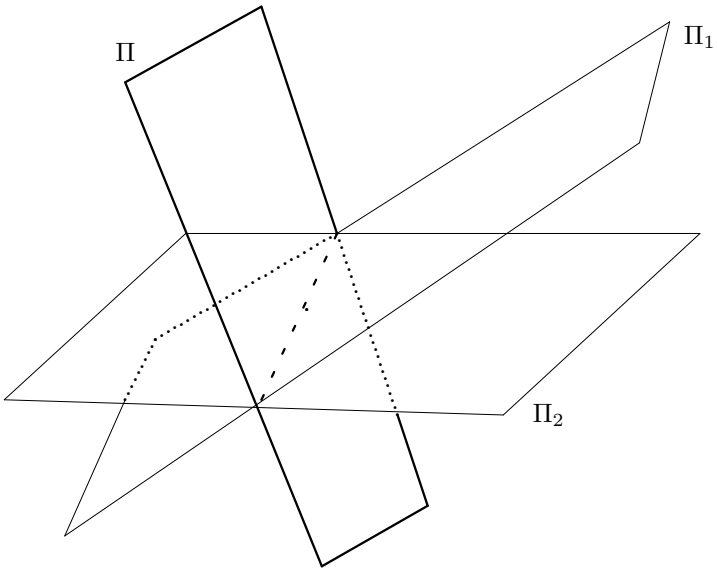


Figure 1.16

1.4 Reflections in a line or a plane

A line or plane can serve as a **mirror** so that a point P is reflected in a point P' in such a way that P and P' are symmetric with respect to the line or plane, as in Figure Figure 1.17 . From the figure the point P' is given by

$$\underline{p}' = \underline{OP'} = \underline{OP} + 2\underline{PQ} = \underline{p} + 2\left(\underline{q} - \underline{p}\right) = 2\underline{q} - \underline{p} \tag{1.26}$$

where Q is the foot of the perpendicular from P to the line or plane and we have written $\underline{p}'$ for the position vector of P'.

Example 1.4.1 *Consider example 1.2.7 of a line in the x-y plane. Using equation (1.15) and equation (1.26), the reflection of $\underline{p} = (p_1, p_2)$ in the line is*

$$\begin{aligned} \underline{p}' &= 2\underline{q} - \underline{p} = 2\left(\frac{1}{4}\left(3p_1 + \sqrt{3}p_2\right), \frac{1}{4}\sqrt{3}p_1 + \frac{1}{4}p_2\right) - (p_1, p_2) \\ &= \left(\frac{1}{2}p_1 + \frac{1}{2}\sqrt{3}p_2, \frac{1}{2}\sqrt{3}p_1 - \frac{1}{2}p_2\right) \end{aligned} \tag{1.27}$$

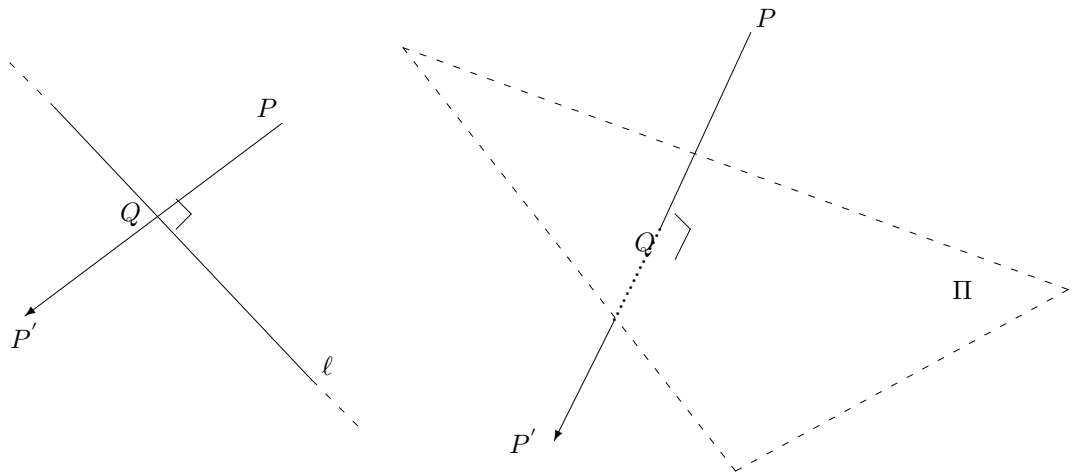
Example 1.4.2 *Consider (1) of Example 1.3.8 where we found the foot of the perpendicular from $\underline{p}$ to the plane $x + 2y + z = -4$ given by equation (1.22):*

$$\underline{q} = (p_1, p_2, p_3) - \frac{4 + p_1 + 2p_2 + p_3}{6}(1, 2, 1)$$

From this it follows that the reflection P' of P in the plane is the point with position

$$\underline{p}' = 2\underline{q} - \underline{p} = \underline{p} - \frac{4 + p_1 + 2p_2 + p_3}{3}(1, 2, 1)$$

Exercise 20 *1. Using equation (1.12) - which finds the foot Q of the perpendicular from P to the line ℓ with parametric equation $\underline{r} = \underline{a} + t\underline{u}$ - find a formula for the reflection P' of P in the line ℓ.*



Reflection in a line ℓ Reflection in a plane Π

Figure 1.17

2. *Using equation (1.24) - which finds the foot Q of the perpendicular from P to the plane with Cartesian equation $\underline{n} \cdot \underline{r} + c = 0$ - find a formula for the reflection P' of P in the given plane.*

3. *Specialize your formulae when $\underline{a} = \underline{0}$ and $c = 0$.*

4. *Using the answers to Questions (2) and (3), or otherwise, find formulae for reflections in the planes $x + 2y + z = 0$, $3x - y - 2z = 1$ and $3x - y - 2z = 0$.*

5. *In the $x - y$ plane find a formula for reflection in the line passing through the origin and making an angle of $60°$ with the positive $x-$axis.*

6. ** In the $x - y$ plane find a formula for reflection in the line passing through the origin and making an angle of θ radians with the positive $x-$axis.*

Hint 21 *All these questions use equation (1.26)*

1.5 Summary of Chapter 1

1.5.1 Points in Space, Vectors

(See section 1.1).
A point $P = (p_1, p_2, p_3)$ in space is identified with its position vector $\underline{OP} = \underline{p}$. The displacement from A to B is $\underline{AB} = \underline{b} - \underline{a}$.

Vectors $\underline{u} = (u_1, u_2, u_3)$ are **underlined**.

Scalars α,β ,...,s, t,... are real numbers.

The product of a vector $\underline{a}$ by a scalar γ is $\gamma\underline{a} = (\gamma a_1, \gamma a_2, \gamma a_3)$. Non-zero vectors $\underline{a}$ and $\underline{b}$ are parallel if $\underline{b} = s\underline{a}$ for a non-zero scalar s.

Vectors $\underline{a}$ and $\underline{b}$ are added:

$$\underline{a} + \underline{b} = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$$

Addition satisfies the triangle law $\underline{AB} + \underline{BC} = \underline{AC}$.

The dot product of vectors $\underline{a}$ and $\underline{b}$ is

$$\underline{a} \cdot \underline{b} = a_1b_1 + a_2b_2 + a_3b_3$$

The length or magnitude of $\underline{a}$ is

$$|\underline{a}| = \sqrt{\underline{a} \cdot \underline{a}}$$

The angle θ between $\underline{a}$ and $\underline{b}$ is given by $\underline{a} \cdot \underline{b} = |\underline{a}|\,|\underline{b}| \cos\theta$.

The unit vector in the direction of $\underline{a} \neq \underline{0}$ is

$$\widehat{\underline{a}} = \frac{1}{|\underline{a}|}\underline{a}$$

The component of $\underline{a}$ in the direction $\underline{b} \neq \underline{0}$ and the projection of $\underline{a}$ on $\underline{b}$ are respectively

$$\underline{a} \cdot \widehat{\underline{b}} \quad \text{and} \quad \left(\underline{a} \cdot \widehat{\underline{b}}\right)\widehat{\underline{b}}$$

1.5.2 Straight lines

(See section 1.2).

A parametric equation of the straight line ℓ through $\underline{a}$ with direction $\underline{u} \neq \underline{0}$ is

$$\underline{r} = \underline{a} + t\underline{u} \qquad (t \in \Re)$$

The foot Q of the perpendicular from P to ℓ is found by solving

$$\left(\underline{a} + t\underline{u} - \underline{p}\right) \cdot \underline{u} = 0$$

for t and then substituting this value back into $\underline{a} + t\underline{u}$.

Lines ℓ_1 and ℓ_1 are parallel if they have parallel directions and are skew if they are not parallel and do not intersect.

1.5.3 Planes

(See section 1.3).

A generic equation of the plane containing two non-parallel vectors $\underline{u}$ and $\underline{v}$ and passing through the point A is

$$\underline{r} = \underline{a} + s\underline{u} + t\underline{v} \qquad (s, t \in \Re)$$

A vector $\underline{n} \neq \underline{0}$ is normal to the plane if it is perpendicular to every line lying in the plane.

A Cartesian equation of the plane has the form

$$(\underline{r} - \underline{a}) \cdot \underline{n} + c = (x_1 - a_1)\, n_1 + (x_2 - a_2)\, n_2 + (x_3 - a_3)\, n_3 + c = 0$$

where c is a constant.

The foot Q of the perpendicular from a point P to the plane is found by solving

$$\underline{n} \cdot \left(\underline{p} + t\underline{n}\right) + c = 0$$

for t and then substituting this into the line $\underline{p} + t\underline{n}$

1.5.4 Projections and reflections

(See equation 1.7, subsection 1.2.5, example 1.3.8 and subsection 1.4).

When a line or plane goes through the origin the foot Q of the perpendicular from P to the line or plane is called the projection of P on the plane.

The reflection P' of a point P in a line or plane is given by

$$\underline{p}' = 2\underline{q} - \underline{p}$$

where Q is the foot of the perpendicular from P to the line or plane.

Chapter 2

Matrices and the Solution of Simultaneous Linear Equations

Linear algebra is used in almost every branch of pure and applied mathematics. It is indispensable for all branches of engineering, the physical sciences (including physics, chemistry and biology), statistics and for operations research (mathematics applied to industry and commerce).

Although most results in this course are valid for various number systems (e.g. rational real or complex numbers), it is assumed that all numbers and number-variables dealt with are **real**, i.e. range over the *real number system* $\Re$. Numbers will also be referred to as **scalars**.

2.1 The Solution of Linear Equations

At school you learned how to solve up to three simultaneous linear equations for three unknowns, say x_1, x_2, x_3, In Chapter 1 we saw how such equations arise in analytic geometry and used vector notation such as a row $\underline{x} = (x_1, x_2, x_3)$ to represent the variables x_1, x_2 and x_3. In this chapter we will learn some systematic techniques for solving m equations for n unknowns x_1, x_2,...,x_n, and we will see the need to carefully distinguish **rows** from **columns**. While geometry often serves as an inspiration for various algebraic results, it is not possible to draw an actual picture representing a system involving n variables when $n \geq 4$.

2.1.1 Some examples

Example 2.1.1 *Find all solutions that satisfy simultaneously*

$$\begin{array}{rcrcrl} 2x_1 & + & 3x_2 & = & -1 & \text{(a)} \\ x_1 & - & 4x_2 & = & 5 & \text{(b)} \end{array}$$

Solution:

1. One way: Make x_1 "the subject of the formula" in (b), i.e. $x_1 = 5 + 4x_2$ and substitute $5 + 4x_2$ for x_1 in (a) to give $2(5 + 4x_2) + 3x_2 = -1$, so $x_2 = -1$ and $x_1 = 1$.
2. We prefer to replace the above system by a sequence of *equivalent systems* (i.e. that have exactly the same solutions), known as **Gauss reduction**.

   (a) (First equivalent system). Multiply (b) by 2 and leave (a) unchanged:

$$\begin{array}{rcrcrll} 2x_1 & + & 3x_2 & = & -1 & \text{leave unchanged} & \text{(a)} \\ 2x_1 & - & 8x_2 & = & 10 & 2 \times \text{(b)} & \text{(b)} \end{array}$$

(b) (Second equivalent system). In the new system subtract (a) from (b), to obtain:

$$\begin{array}{rcrcrlll} 2x_1 & + & 3x_2 & = & -1 & \text{leave unchanged} & \text{(a)} \\ & - & 11x_2 & = & 11 & [\,(\text{-}1) \times \text{(a)}]+\text{(b)} & \text{(b)} \end{array}$$

It is now obvious from (b) that $x_2 = -1$. By back-substitution in (a) we get $x_1 = 1$. For reasons that will be become clear later we write this **unique** solution as a *column* vector

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Example 2.1.2 *Consider*

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & 3x_3 & = & -1 & \text{(a)} \\ x_1 & + & x_2 & + & 4x_3 & = & -3 & \text{(b)} \\ -x_1 & + & x_2 & - & 2x_3 & = & 1 & \text{(c)} \end{array} \tag{2.1}$$

Solution:

$$\begin{array}{rcrcrcrll} x_1 & + & x_2 & + & 3x_3 & = & -1 & \text{leave unchanged} & \text{(a)} \\ & & & & x_3 & = & -2 & [(-1) \times \text{(a)}]+\text{(b)} & \text{(b)} \\ -x_1 & + & x_2 & - & 2x_3 & = & 1 & \text{leave unchanged} & \text{(c)} \end{array}$$

$$\begin{array}{rcrcrcrll} x_1 & + & x_2 & + & 3x_3 & = & -1 & \text{leave unchanged} & \text{(a)} \\ & & & & x_3 & = & -2 & \text{leave unchanged} & \text{(b)} \\ & & 2x_2 & + & x_3 & = & 0 & \text{(a) + (c)} & \text{(c)} \end{array}$$

Finally, interchange the equations (b) and (c):

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & 3x_3 & = & -1 & \text{(a)} \\ & & 2x_2 & + & x_3 & = & 0 & \text{(b)} \\ & & & & x_3 & = & -2 & \text{(c)} \end{array} \tag{2.2}$$

Because the coefficients of x_1 in (b) and (c) are zero and also because the coefficient of x_2 in (c) is zero, we say the system is in (row) **echelon** or (upper) **triangular** form. From (c), $x_3 = -2$. Back-substitution into (b) gives $x_2 = 1$. Finally, back-substitution of $x_3 = -2$ and $x_2 = 1$ into (a) gives $x_1 = 4$. The solution is

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix} \tag{2.3}$$

Example 2.1.3 *Find all solutions $\underline{x}$ to the system*

$$\begin{array}{rcrcrcrcrl} x_1 & + & x_2 & + & 3x_3 & + & 2x_4 & = & -1 & \text{(a)} \\ & & 2x_2 & + & x_3 & + & x_4 & = & 0 & \text{(b)} \\ & & & & x_3 & - & x_4 & = & -2 & \text{(c)} \end{array} \tag{2.4}$$

Solution:

There are only *three* equations for *four* unknowns and the equations are already in echelon form. One of the unknowns (say x_4) can be **any** number and back-substitution gives in turn $x_3 = x_4 - 2$, $x_2 = -\frac{1}{2}(x_3 + x_4) = 1 - x_4$ and $x_1 = -1 - x_1 - x_2 - 3x_3 - 2x_4 = 4 - 4x_4$. It is customary to let $x_4 = t$, so

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4-4t \\ 1-t \\ t-2 \\ t \end{bmatrix} \tag{2.5}$$

is the *general* solution. The **parameter** $t \in \Re$ can be any real number.

Remark 22 *Here we have* 3 *equations for* 4 *unknowns and the solution shows that we have an infinity of solutions. In Example 1.3.2 of Chapter 1, we found a vector* $\underline{n} \neq \underline{0}$ *normal to a plane. For* $\beta \neq 0$, $\beta\underline{n}$ *is also a normal, so again we have an infinite number of solutions, this time of two equations in three unknowns. Example 1.3.4 of Chapter 1 involved solving one equation for three unknowns and this involved two parameters.*

These examples and also Example 2.1.3 illustrate a general result:
If a system of equations has more unknowns than equations and there is a solution, then the system has infinitely many solutions. See, for example, 2.2.15 below. In Exercise 52, No.5 you will get a more complete picture of this theorem.

Example 2.1.4 *For which values of the variables does the following system have a solution?*

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & 3x_3 & = & -1 & \text{(a)} \\ x_1 & + & 3x_2 & + & 4x_3 & = & -1 & \text{(b)} \\ 2x_1 & + & 4x_2 & + & 7x_3 & = & 5 & \text{(c)} \end{array}$$

Solution:

$$\begin{array}{rcrcrcrll} x_1 & + & x_2 & + & 3x_3 & = & -1 & \text{leave unchanged} & \text{(a)} \\ & & 2x_2 & + & x_3 & = & 0 & [(-1)\times\text{(a)}]+\text{(b)} & \text{(b)} \\ & & 2x_2 & + & x_3 & = & 7 & [(-2)\times\text{(a)}]+\text{(c)} & \text{(c)} \end{array}$$

The system cannot have a solution since (b) and (c) are contradictory. The system is **inconsistent** (or the equations are **incompatible**) and the solution set is empty.

2.2 Matrices

A rectangular array A of numbers with m rows and n columns is called an $m \times n$ ("m by n") **matrix**. Matrices will form a new kind of algebra and it is customary to place brackets (we will use square ones) around such arrays.

2.2.1 Examples of matrices

$\begin{bmatrix} 5 & -1 & 13 & -1 \\ -1 & 1 & -12 & \frac{1}{3} \\ 1 & 0 & 40 & -3 \end{bmatrix}$ is a 3×4 matrix, $\begin{bmatrix} -\frac{1}{2} & 21 & -3 \\ -1 & 5 & 0 \\ 0 & -2 & 7 \\ 1 & -2 & 6 \end{bmatrix}$ is a 4×3 matrix.

$\begin{bmatrix} 2 & \frac{3}{4} & -1 \\ 1 & -6 & -\frac{2}{5} \end{bmatrix}$ is a 2×3 matrix; $\begin{bmatrix} 3 & 2 & \frac{1}{2} \\ -1 & 0 & 5 \\ 7 & 8 & -2 \end{bmatrix}$ is a 3×3 matrix. The column $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is a 3×1 matrix, while the row $\underline{y} = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \end{bmatrix}$ is a 1×4 matrix.

One speaks of the *row* entries of a column and the *column* entries of a row.

2.2.2 Standard notation for an $m \times n$ matrix A

Given that A is an $m \times n$ matrix, the standard notation is to write $A = [a_{ij}]$.
Here a_{ij} is the entry of A in the i^{th} row and j^{th} column. One calls a_{ij} the (i, j) entry of A. The first index i labels the rows and has the range $i = 1, \cdots, m$, while the second index j labels the columns and has the range $j = 1, \cdots, n$.
For example, if $m = 4$ and $n = 5$,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix} \tag{2.6}$$

is a 4×5 matrix. One reads a_{12} as "a one-two", not as "a twelve". Likewise, a_{34} is "a three-four", not "a thirty four". For example, consider the 2×3 matrix

$$A = \begin{bmatrix} 2 & 3 & -1 \\ 1 & -4 & 5 \end{bmatrix}$$

The $(1, 1)$ entry is $a_{11} = 2$. It is the first entry in column 1 and also the first entry in row 1. The $(2, 1)$ entry $a_{21} = 1$ is the second row entry of column 1 and the first entry in row 2. The numbers $a_{12} = 3$, $a_{22} = -4$, are the first and second entries of column 2 respectively. Row 2 has entries $a_{21} = 1$, $a_{22} = -4$, $a_{23} = 5$ respectively. The $(1, 3)$ entry of A is $a_{13} = -1$ the $(2, 3)$ entry is $a_{23} = 5$.

It is important to keep in mind that a_{ij} is in **row** i and in **column** j of A.

Sometimes we write $A = A_{m\times n}$ to indicate that A is $m \times n$. Thus if A has two rows and three columns, we may write $A = A_{2\times 3}$.

Exercise 23 *1. For each matrix A in 2.2.1 write down*

- *(a) each of its entries in the form $a_{ij} = ...$*
- *(b) each of its rows and columns. What notation do you think would be suitable for denoting row i of a matrixA? What notation do you think would be suitable for denoting column j of a matrix A?*
  In Chapter 3, 3.1.1, we will develop a systematic notation for these; here you are only asked to use your imagination.

2. Find a formula for the entries a_{ij} in the following 4×4 matrix

$$A = \begin{bmatrix} 2 & -3 & 4 & -5 \\ -3 & 4 & -5 & 6 \\ 4 & -5 & 6 & -7 \\ -5 & 6 & -7 & 8 \end{bmatrix}$$

3. The $n \times n$ Hilbert matrix $H_n = [h_{ij}]$ is defined by

$$h_{ij} = \frac{1}{i + j - 1} \quad (i = 1, \ldots, n; \quad j = 1, \ldots, n)$$

Write down the matrices H_1, H_2, H_3, H_4.

2.2.3 Expressing m equations in n unknowns as a matrix equation $A\underline{x} = \underline{b}$

A system of linear equations has a matrix of coefficients, or **coefficient matrix**. For example, the coefficient matrix of Example 2.1.4 is

$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & 4 \\ 2 & 4 & 7 \end{bmatrix}$$

If we are given m simultaneous equations in n unknowns x_1, x_2, ... , x_n, there is a standard way to express the equations as a single matrix equation.

In the case of Example 2.1.4 this expression is

$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & 4 \\ 2 & 4 & 7 \end{bmatrix} \underline{x} = \begin{bmatrix} -1 \\ -1 \\ 5 \end{bmatrix} \tag{2.7}$$

The system of Example 2.1.2 (equations (2.1)) has for coefficient matrix

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 4 \\ -1 & 1 & -2 \end{bmatrix}$$

The standard corresponding matrix equation is

$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 4 \\ -1 & 1 & -2 \end{bmatrix} \underline{x} = \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix} \tag{2.8}$$

What this means is **the equality of two column matrices**:

$$\begin{bmatrix} x_1 + x_2 + 3x_3 \\ x_1 + x_2 + 4x_3 \\ -x_1 + x_2 - 2x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix} \tag{2.9}$$

The system of equations of Example 2.1.3 has the matrix expression

$$\begin{bmatrix} 1 & 1 & 3 & 2 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix} \tag{2.10}$$

Using the general 4×5 matrix of (2.6), the following matrix equation expresses a general system of 4 equations in 5 unknowns x_1, x_2, ... , x_5:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \tag{2.11}$$

Written out, we again have the equality of two column matrices:

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \tag{2.12}$$

In general, let A be the matrix of coefficients of m simultaneous equations in n unknowns x_1, x_2, ..., x_n. If the right-hand side of equation i is b_i, the standard way to express the equations in matrix form is

$$A\underline{x} = \underline{b} \tag{2.13}$$

Here the unknowns x_j are written as an $n \times 1$ column matrix (a **column vector**) $\underline{x}$ and $\underline{b}$ is the $m \times 1$ column of the b_i.

Observe that for the matrix equation (2.13) to be meaningful the number of rows of A must be the same as the number of rows of $\underline{b}$ and the number of columns of A must equal the number of rows of $\underline{x}$.

2.2.4 The dot product generalized

Consider the equations of the left-hand side of Example 2.1.2 or of equation (2.9). Each entry is a dot product, e.g. the third entry is $(-1,1,-2)\cdot(x_1,x_2,x_3) = -x_1 + x_2 - 2x_3$. In the next chapter we will emphasize this as the **product of a row and a column**:

$$\begin{bmatrix} -1 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -x_1 + x_2 - 2x_3$$

Likewise, the left-hand sides of the first and second equations are respectively

$$\begin{bmatrix} 1 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 + x_2 + 3x_3$$

$$\begin{bmatrix} 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 + x_2 + 4x_3$$

Similarly, the left-hand side of the first equation of Example 2.1.3 is the product

$$\begin{bmatrix} 1 & 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 + x_2 + 3x_3 + 2x_4$$

The left-hand side of equation (2.10) in Example 2.1.3 is the column with three products

$$\begin{bmatrix} \begin{bmatrix} 1 & 1 & 3 & 2 \end{bmatrix} \underline{x} \\ \begin{bmatrix} 0 & 2 & 1 & 1 \end{bmatrix} \underline{x} \\ \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} \underline{x} \end{bmatrix} = \begin{bmatrix} x_1 + x_2 + 3x_3 + 2x_4 \\ 2x_2 + x_3 + x_4 \\ x_3 - x_4 \end{bmatrix}$$

Remark 24 *The last examples show that we are dealing with a product which is a **generalization** of the dot product of Chapter 1 (subsection 1.1.12).*
*As remarked earlier, we will later develop special notations for row i and column j of A. What is more, we will emphasize $A\underline{x}$ as the **product** of the matrices A and $\underline{x}$. See 3.1.6.*

2.2.5 Detached Coefficients

We can also express the above problems compactly by considering the array formed by attaching the right- hand side $\underline{b}$ to the coefficient matrix A in the form $[A, \underline{b}]$. Corresponding to the problem of solving the matrix equation (2.13) we call $R = [A, \underline{b}]$ the **array of detached coefficients** or **the matrix A augmented by $\underline{b}$**. Operations done on the equations have their exact counterpart when done on the rows of R.

1. Example 2.1.1 above can be expressed as the matrix equation

$$\begin{bmatrix} 2 & 3 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \end{bmatrix} \tag{2.14}$$

   or as the array of detached coefficients:

$$\begin{matrix} 2 & 3 & -1 \\ 1 & -4 & 5 \end{matrix} \tag{2.15}$$

   We have omitted brackets around the array. Here x_1 and x_2 can be thought of as labels of the first two columns of the array. The third column of (2.15) is the right-hand side of

(2.14). It is clear that two equations in two unknowns determine and are determined by such an array. Instead of performing operations on the given equations, we perform the same operations on this array.
In the case of Example 2.1.1 these steps are:

$$\begin{array}{cccl} 2 & 3 & -1 & \\ 2 & -8 & 10 & 2R_2 \text{ (multiply Row 2 by 2)} \end{array}$$

$$\begin{array}{cccl} 2 & 3 & -1 & \\ 0 & -11 & 11 & -R_1 + R_2 \text{ (add } (-1) \times \text{Row 1 to Row 2)} \end{array}$$

From the last array we read off $-11x_2 = 11$, so $x_2 = -1$. The first row reads $2x_1 + 3x_2 = -1$ and $x_1 = 1$. This is the same solution we found before.

2. Example 2.1.2 has an expression as the matrix equation (2.8) and as an array of detached coefficients the equations are:

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 1 & 1 & 4 & -3 \\ -1 & 1 & -2 & 1 \end{array}$$

Here x_1, x_2, and x_3 label the first three columns. The steps leading to the solution are:

$$\begin{array}{cccll} 1 & 1 & 3 & -1 & \\ 0 & 0 & 1 & -2 & -R_1 + R_2 \\ -1 & 1 & -2 & 1 & \end{array}$$

$$\begin{array}{cccll} 1 & 1 & 3 & -1 & \\ 0 & 0 & 1 & -2 & \\ 0 & 2 & 1 & 0 & R_1 + R_3 \end{array}$$

Finally, we interchange rows 2 and 3:

$$\begin{array}{cccll} 1 & 1 & 3 & -1 & \\ 0 & 2 & 1 & 0 & R_2 \leftrightarrow R_3 \\ 0 & 0 & 1 & -2 & \end{array} \tag{2.16}$$

The equations are in echelon form and we can read off the solution:
$x_3 = -2$, $2x_2 + x_3 = 0$, so $x_2 = 1$ and finally from $x_1 + x_2 + 3x_3 = -1$ we get $x_1 = 4$. We have

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix}$$

This is equation (2.3), found earlier.

Observation: It is important to observe that the right-hand side of the original set of equations is a **linear combination** of the columns of the coefficient matrix with coefficients x_1, x_2 and x_3:

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} 4 + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix} (-2) = \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix} \tag{2.17}$$

Linear combinations were introduced for row vectors in Chapter 1, subsection 1.1.18 and will be systematically developed in Chapter 3. See subsection 3.1.2 and especially 3.1.9.

Note that it is natural to write the coefficients on the right of the column vectors since $\underline{x}$ appears on the right of A when we express the equations in the form $A\underline{x} = \underline{b}$.

3. Example 2.1.3 (equations (2.4)) has the matrix expression (2.10). The corresponding array of detached coefficients is already in echelon form:

$$\begin{array}{ccccc} 1 & 1 & 3 & 2 & -1 \\ 0 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 & -2 \end{array} \tag{2.18}$$

The general solution can be read off as before by putting $x_4 = t$ and solving in turn for x_3, x_2 and x_1:

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4-4t \\ 1-t \\ t-2 \\ t \end{bmatrix}$$

This is just the solution (2.5) found earlier. A particular solution is found by giving the parameter t some value, eg. $t = 3$. In that case $x_1 = -8$, $x_2 = -2$, $x_3 = 1$ and $x_4 = 3$. As a linear combination we have

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}(-8) + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}(-2) + \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} 3 = \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix} \tag{2.19}$$

4. Example 2.1.4 has two expressions, one as equation (2.7), and also as the array of detached coefficients (augmented matrix)

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 1 & 3 & 4 & -1 \\ 2 & 4 & 7 & 5 \end{array}$$

The steps leading to the solution are:

$$\begin{array}{cccc} 1 & 1 & 3 & -1 & \\ 0 & 2 & 1 & 0 & -R_1 + R_2 \\ 0 & 2 & 1 & 7 & -2R_1 + R_3 \end{array}$$

Taking it a step further, we obtain the echelon (triangular) form:

$$\begin{array}{cccc} 1 & 1 & 3 & -1 & \\ 0 & 2 & 1 & 0 & \\ 0 & 0 & 0 & 7 & -R_2 + R_3 \end{array} \tag{2.20}$$

The last row reads $0x_1 + 0x_2 + 0x_3 = 7$, an impossibility. The given equations are inconsistent (incompatible); they do not have a solution.

5. As a detached array, the equations (2.11) appear as

$$\begin{array}{cccccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & b_3 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_4 \end{array} \tag{2.21}$$

Remark 25 *Our convention is that an operation performed on row i is written next to row i* ***after*** *the operation has been done.*

2.2.6 Permissible operations on equations: Elementary row operations. Equivalent row operations on the augmented matrix $R = [A, \underline{b}]$

The row operations which do not alter the solution set of a system $A\underline{x} = \underline{b}$ of m simultaneous linear equations for n unknowns are:

1. Interchange equations i and j (interchange rows R_i and R_j of the corresponding array of detached coefficients).
2. Multiply equation i (multiply row R_i of the corresponding array) by a scalar $\beta \neq 0$.
3. For any scalar γ and $i \neq j$, add γ times equation i to equation j (add γ times row R_i of the corresponding array to row R_j).

The above three operations done on an array are called **elementary row operations** (abbreviated as **eros**). Note that forming βR_i is just like multiplying a vector in 3-space by a scalar, except that now the vector has $n+1$ entries. The type (3) operation replaces R_j with $\gamma R_i + R_j$, where addition of rows is addition of vectors.

What is obvious (and assumed at school) is that if $\underline{x}$ is a simultaneous solution of the m given equations **before** any one of the above eros is performed, then $\underline{x}$ remains a solution **after** the operation is performed.

To see that the reverse is true, suppose we have just performed the elementary row operation of type (3) and that $\underline{x}$ is a solution to the new system of equations. Adding $-\gamma$ times equation i ($-\gamma$ times row R_i of the augmented array) to equation j (row R_j of the array) in the new system of equations (array) brings us back to the system before the type (3) operation was performed. This shows that $\underline{x}$ is also a solution to the old system.

For example, consider example 2.1.2 again and the first step toward a solution, as in 2:

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 1 & 1 & 4 & -3 \\ -1 & 1 & -2 & 1 \end{array}$$

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 0 & 0 & 1 & -2 \\ -1 & 1 & -2 & 1 \end{array} \qquad -R_1 + R_2$$

Adding row 1 to row 2 of this array restores the original array:

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 1 & 1 & 4 & -3 \\ -1 & 1 & -2 & 1 \end{array} \qquad R_1 + R_2$$

The following more general setting refers to the array (2.21). Suppose we add γ times Row 2 to Row 3, obtaining

$$\begin{array}{cccccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & b_2 \\ \gamma a_{21} + a_{31} & \gamma a_{22} + a_{32} & \gamma a_{23} + a_{33} & \gamma a_{24} + a_{34} & \gamma a_{25} + a_{35} & \gamma b_2 + b_3 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_4 \end{array} \qquad \gamma R_2 + R_3$$

Then,

$$\begin{array}{cccccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & b_3 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_4 \end{array} \qquad -\gamma R_2 + R_3$$

brings us back to the original array.

Remark 26 *For convenience we have been using R_i for row i of an array R. In the next chapter we will use a slightly different notation for row i of a matrix.*

2.2.7 Arrays in row echelon form (row EF): Proper Definition

An array is in row echelon form if

1. All zero rows come after the non-zero rows.

2. For non-zero rows, the first non-zero entry in row $i+1$ comes **later** than the first non-zero entry in row i.

Remark 27 *Row echelon forms are also called upper triangular. In subsection 3.4.14 of Chapter 3 we will also discuss elementary column operations and column echelon forms. There, the distinction between the row and column forms will matter.*

2.2.8 How an array gets reduced to row echelon form

At the start, and at any later stage, we arrange the rows so that

a all zero rows come after non-zero rows and

b for the non-zero rows, the first non-zero entry in row $i+1$ does not come **earlier** than the first non-zero entry in row i.

Suppose that only the first r rows are non-zero and in row echelon form. The first non-zero entry in row r is called the **pivot element** and row r is the **pivot** or **reference** row.
We assume that (a) and (b) hold. If the remaining rows are zero we are finished. Otherwise, multiples of row r are added to rows **below** it in such a way that all entries below the pivot element become zeros. The first $r+1$ rows are then in echelon form and this remains the case after arranging rows $r+1$, $r+2$,... so that (a) and (b) hold.
The process ends when the whole array is in row echelon form and then each non-zero row has a pivot element.

The columns of the array containing a pivot element are called **pivot columns**. (This will be used in 2.2.15 below and also in subsection 3.4.9).

2.2.9 Further examples

Example 2.2.1 *Reduce to row echelon form*

$$\begin{matrix} -3 & -10 & 2 \\ 7 & 5 & 6 \end{matrix}$$

Solution (method 1)

$$\begin{matrix} 1 & \frac{10}{3} & -\frac{2}{3} & \left(-\frac{1}{3}R_1\right) \\ 7 & 5 & 6 & \end{matrix}$$

$$\begin{matrix} 1 & \frac{10}{3} & -\frac{2}{3} & \text{ref row, pivot1} \\ 0 & -\frac{55}{3} & \frac{32}{3} & -7R_1 + R_2 \end{matrix}$$

Remark 28 *This illustrates how the pivot element can always be made* 1.

Method 2 (Fraction-free solution)

$$\begin{array}{cccc} -21 & -70 & 14 & 7R_1 \quad \text{ref row} \\ 21 & 15 & 18 & 3R_2 \end{array}$$

$$\begin{array}{cccc} -21 & -70 & 14 & \\ 0 & -55 & 32 & R_1 + R_2 \end{array}$$

Remark 29 *This example illustrates how it is possible to obtain a fraction-free answer when all entries are integers or even fractions.*

Example 2.2.2 *Reduce the array of detached coefficients to row echelon form and hence solve*

$$\begin{bmatrix} 1 & 0 & 2 \\ 3 & 0 & 5 \\ -5 & 1 & -10 \end{bmatrix} \underline{x} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

Solution:

The augmented matrix (without brackets) is

$$\begin{array}{cccc} 1 & 0 & 2 & -1 \quad \text{ref row} \\ 3 & 0 & 5 & 1 \\ -5 & 1 & -10 & 2 \end{array}$$

Elementary row operations (eros) bringing the detached array into echelon form are:

$$\begin{array}{cccc} 1 & 0 & 2 & -1 \\ 0 & 0 & -1 & 4 \\ 0 & 1 & 0 & -3 \end{array} \quad \begin{array}{l} \\ -3R_1 + R_2 \\ 5R_1 + R_3 \end{array} \quad \text{then} \quad \begin{array}{cccc} 1 & 0 & 2 & -1 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & -1 & 4 \end{array} \quad R_2 \leftrightarrow R_3$$

From this row echelon form we find the unique solution

$$\underline{x} = \begin{bmatrix} 7 \\ -3 \\ -4 \end{bmatrix}$$

Example 2.2.3 *Triangulate and hence solve*

$$\begin{bmatrix} 3 & -3 & 1 & 10 \\ 2 & -3 & -1 & 3 \\ -1 & 2 & 0 & 3 \\ 1 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -6 \\ 4 \end{bmatrix} \tag{2.22}$$

Solution:

The initial array of detached coefficients is

$$\begin{array}{ccccc} 3 & -3 & 1 & 10 & 2 \\ 2 & -3 & -1 & 3 & 0 \\ -1 & 2 & 0 & 3 & -6 \\ 1 & -1 & 1 & 2 & 4 \end{array}$$

The steps giving equivalent systems are:

Step 1:

$$\begin{array}{ccccc} 1 & -1 & 1 & 2 & 4 \\ 2 & -3 & -1 & 3 & 0 \\ -1 & 2 & 0 & 3 & -6 \\ 3 & -3 & 1 & 10 & 2 \end{array} \quad R_1 \leftrightarrow R_4 \ \text{ interchange row 1 and row 4} \quad \text{ref row}$$

We have interchanged rows 1 and 4 as it is convenient to have a 1 as pivot element.
Step 2

$$\begin{array}{ccccc} 1 & -1 & 1 & 2 & 4 \\ 0 & -1 & -3 & -1 & -8 \\ 0 & 1 & 1 & 5 & -2 \\ 0 & 0 & -2 & 4 & 10 \end{array} \quad \begin{array}{c} \\ -2R_1 + R_2 \quad \text{new ref row} \\ R_1 + R_3 \\ (-3)\, R_1 + R_4 \end{array}$$

At this point the *first two* rows are in row echelon form (row EF).
Step 3. Use row 2 for reference row to obtain

$$\begin{array}{ccccc} 1 & -1 & 1 & 2 & 4 \\ 0 & -1 & -3 & -1 & -8 \\ 0 & 0 & -2 & 4 & -10 \\ 0 & 0 & -2 & 4 & -10 \end{array} \quad \begin{array}{c} \\ \\ R_2 + R_3 \quad \text{new ref row} \\ \end{array}$$

Step 4. Row 3 is the new reference row:

$$\begin{array}{ccccc} 1 & -1 & 1 & 2 & 4 \\ 0 & -1 & -3 & -1 & -8 \\ 0 & 0 & -2 & 4 & -10 \\ 0 & 0 & 0 & 0 & 0 \end{array} \quad \begin{array}{c} \\ \\ \\ -\, R_3 + R_4 \end{array}$$

This is in echelon form but we can simplify still further:

$$\begin{array}{ccccc} 1 & -1 & 1 & 2 & 4 \\ 0 & 1 & 3 & 1 & 8 \\ 0 & 0 & 1 & -2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array} \quad \begin{array}{c} \\ -R_2 \\ \frac{1}{-2} R_3 \\ \end{array} \tag{2.23}$$

The required general solution to Example 2.2.3 is with $x_4 = t$ the parameter, $x_3 = 2t + 5$, $x_2 = -7t - 7$ and $x_1 = -11t - 8$. Or,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -11t - 8 \\ -7t - 7 \\ 2t + 5 \\ t \end{bmatrix} \tag{2.24}$$

As a linear combination of the columns of the coefficient matrix,

$$\begin{bmatrix} 3 \\ 2 \\ -1 \\ 1 \end{bmatrix} (-11t - 8) + \begin{bmatrix} -3 \\ -3 \\ 2 \\ -1 \end{bmatrix} (-7t - 7) + \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix} (2t + 5) + \begin{bmatrix} 10 \\ 3 \\ 3 \\ 2 \end{bmatrix} t = \begin{bmatrix} 2 \\ 0 \\ -6 \\ 4 \end{bmatrix} \tag{2.25}$$

Remark 30 *Each column vector has 4 entries instead of 3 and so we cannot visualize these vectors in space. Nevertheless, this should not be a problem. In Chapter 3 we will consider $n-$ dimensional vectors. See subsections 3.1.2 3.1.9.*

Remark 31 *Although we associate arrays with simultaneous equations, it is important to realize that such operations can be done on any arrays, i.e. matrices, without any reference to solving equations in the usual sense. Note however the following:*
Reducing a matrix A to to row echelon form amounts to reducing the augmented matrix $[A, \underline{0}]$ to row echelon form. This in turn amounts to solving $A\underline{x} = \underline{0}$. See Remark 34.

Example 2.2.4 *Reduce the following matrix to row echelon form:*

$$A = \begin{bmatrix} 0 & 0 & -3 & -3 & -1 \\ 2 & -2 & -2 & -3 & -2 \\ 0 & 0 & 3 & -2 & -2 \\ 2 & -3 & -4 & 3 & -3 \end{bmatrix}$$

Solution:

In order for (b) in 2.2.8 to hold, first interchange rows 1 and 4 of A:

$$\begin{array}{ccccc} 2 & -3 & -4 & 3 & -3 \\ 2 & -2 & -2 & -3 & -2 \\ 0 & 0 & 3 & -2 & -2 \\ 0 & 0 & -3 & -3 & -1 \end{array} \qquad R_1 \leftrightarrow R_4 \quad \text{ref row}$$

Next:

$$\begin{array}{ccccc} 2 & -3 & -4 & 3 & -3 \\ 0 & 1 & 2 & -6 & 1 \\ 0 & 0 & 3 & -2 & -2 \\ 0 & 0 & -3 & -3 & -1 \end{array} \qquad \begin{array}{l} \\ -R_1 + R_2 \\ \text{new ref row} \\ \end{array}$$

Finally,

$$\begin{array}{ccccc} 2 & -3 & -4 & 3 & -3 \\ 0 & 1 & 2 & -6 & 1 \\ 0 & 0 & 3 & -2 & -2 \\ 0 & 0 & 0 & -5 & -3 \end{array} \qquad \begin{array}{l} \\ \\ \\ R_3 + R_4 \end{array} \tag{2.26}$$

The matrix is now in row echelon form.

Exercise 32 *1. Express the following simultaneous equations in the form $A\underline{x} = \underline{b}$ and use Gauss-reduction and detached coefficients to find solutions. Display the final array in row echelon form. Also, express $\underline{b}$ as a linear combination of the columns of A whenever this is possible.*

(a) i.

$$\begin{array}{rcrcl} x_1 & - & 2x_2 & = & -1 \\ 3x_1 & - & x_2 & = & 1 \end{array}$$

ii.

$$\begin{array}{rcrcl} x & - & 2y & = & -3 \\ 2x & - & 4y & = & -1 \end{array}$$

iii.

$$\begin{array}{rcrcl} x & - & 2y & = & -3 \\ 2x & - & 4y & = & -6 \end{array}$$

iv.

$$\begin{array}{rcrcl} x_1 & - & 2x_2 & = & -3 \\ 3x_1 & - & x_2 & = & 1 \\ 4x_1 & - & 3x_2 & = & 5 \end{array}$$

v.

$$\begin{array}{rcrcl} x & - & 2y & = & -3 \\ 2x & + & y & = & 4 \\ 3x & - & y & = & 1 \end{array}$$

vi.

$$\begin{array}{rcrcl} 0x & + & 0y & = & 0 \end{array}$$

vii.

$$\begin{array}{rcrcl} 3x & + & 0y & = & -3 \end{array}$$

(b) i.

$$\begin{array}{rcrcrcl} x_1 & - & 2x_2 & - & x_3 & = & 0 \\ & & 3x_2 & + & x_3 & = & 1 \end{array}$$

ii.

$$\begin{array}{rcrcrcrcl} x_1 & + & x_2 & + & x_3 & - & x_4 & = & 1 \\ & & 2x_2 & - & x_3 & + & 4x_4 & = & -1 \\ & & & & & & x_4 & = & -2 \end{array}$$

iii.

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & x_3 & = & -1 \end{array}$$

iv.

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & x_3 & - & 5x_4 & = & 0 \end{array}$$

2. *(a) In solving simultaneous equations, if we drop an equation the resulting solution set contains and is usually larger than the original solution set. Illustrate this with an example.*

   (b) Rule 3 of permissable row operations in 2.2.6 says: For any scalar γ and $i \neq j$, add γ times row R_i of the array to row R_j. Why must we have $i \neq j$?

   (c) In an array with rows $R_1, \ldots, R_m$, let $i \neq j$. Suppose that at least one of the numbers α, β is not zero and form the row $\alpha R_i + \beta R_j$. Under what conditions can this new row replace row i or row j? Describe what is going on in terms of the strict rules of 2.2.6.

3. ** (Simplified Elementary Row Operations). Consider the following simplified set of two elementary row operations that can be applied to an array. Show that each row operation in 2.2.6 can be obtained by applying a sequence of these two operations. In other words, the two row operations are* ***equivalent*** *to those in 2.2.6.*

   (a) Multiply row R_i of the array by a scalar $\beta \neq 0$.

   (b) For $i \neq j$ add row R_i to row R_j.

   Hint 33 *In order to add αR_i to R_j when $i \neq j$, we may obviously assume $\alpha \neq 0$. Perform (a) with $\beta = \alpha$ then do (b), i.e. add the resulting row i to row j. What must be done next?*
   It still remains to show that we can interchange rows R_i and R_j of the array using operations like (a) and (b). Try adding row i to row j, then changing the sign of row i.

2.2.10 Homogeneous Equations

A system $A\underline{x} = \underline{b}$ of m equations in n unknowns is called **homogeneous** if in the right-hand side all $b_i = 0$. Such a system always has at least one solution, namely $\underline{x} = \underline{0}$, the column vector with all n entries $x_j = 0$. This is known as the **trivial** solution. The existence of non-trivial solutions is the subject of subsection 2.2.15.
Every system defines a corresponding homogeneous system.

Example 2.2.5 *Consider the following homogeneous system of equations:*

$$\begin{bmatrix} 3 & -3 & 1 & 10 \\ 2 & -3 & -1 & 3 \\ -1 & 2 & 0 & 3 \\ 1 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.27}$$

This is the homogeneous system corresponding to Example 2.2.3. We found the row echelon form of the detached coefficients for that example as (2.23). The row EF for the homogeneous system (2.27) is therefore

$$\begin{array}{ccccc} 1 & -1 & 1 & 2 & 0 \\ 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

The general solution of (2.27) is with $x_4 = t \in \Re$ any real number,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -11t \\ -7t \\ 2t \\ t \end{bmatrix} \tag{2.28}$$

Remark 34 *When solving a homogeneous system $A\underline{x} = \underline{0}$, the column $\underline{0}$ can simply be ignored in the array of detached coefficients.*

Exercise 35 *1. Find a row echelon form of the following matrices:*

(a)

$$A = \begin{bmatrix} 1 & 2 & -3 & -1 \\ -3 & 4 & 1 & 2 \\ -2 & 6 & -2 & 1 \\ 4 & -2 & -4 & -3 \end{bmatrix}$$

(b)

$$B = \begin{bmatrix} 1 & -3 & -2 & 4 \\ 2 & 4 & 6 & -2 \\ -3 & 1 & -2 & -4 \\ -1 & 2 & 1 & -3 \end{bmatrix}$$

(c) How are the matrices A and B related?

2. Let:

$$A = \begin{bmatrix} 1 & 1 & 2 & 1 & -3 \\ 2 & 1 & 5 & 2 & -7 \\ 0 & -1 & 3 & 5 & -4 \\ -2 & 1 & -3 & 9 & 3 \end{bmatrix}$$

(a) Reduce A to echelon form.

(b) Solve the homogeneous system $A\underline{x} = \underline{0}$. Express $\underline{0}$ as a linear combination of the columns of A in a non-trivial way.

Hint 36 *For coefficients take the values of x_i in a homogeneous solution that are not all zero. (The trivial linear combination is the one with all coefficients 0.)*

*The columns of the above A are said to be **linearly dependent**.*

Linear independence and dependence will be studied formally in Chapter 3, section 3.4. See also the theorem 2.2.15 below.

3. (a) Reduce the array of detached coefficients to echelon form and solve

$$\begin{bmatrix} 1 & -1 & 1 & 2 \\ 2 & -2 & -1 & 3 \\ -1 & 1 & 0 & 3 \\ 1 & -2 & 2 & 5 \end{bmatrix} \underline{x} = \begin{bmatrix} 4 \\ 0 \\ -6 \\ 5 \end{bmatrix}$$

(b) *Show that whatever the RHS* $\underline{b}$ *may be, the system always has a unique solution.*

4. (a) *Solve for* $\underline{x}$*:*

$$\begin{bmatrix} 1 & 1 & 2 & 1 \\ 2 & 1 & 5 & 2 \\ 0 & -1 & 3 & 5 \\ 3 & 1 & 10 & 8 \end{bmatrix} \underline{x} = \begin{bmatrix} -3 \\ -7 \\ -4 \\ -14 \end{bmatrix}$$

(b) *Solve the corresponding homogeneous system for* $\underline{x}$*:*

$$\begin{bmatrix} 1 & 1 & 2 & 1 \\ 2 & 1 & 5 & 2 \\ 0 & -1 & 3 & 5 \\ 3 & 1 & 10 & 8 \end{bmatrix} \underline{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

5. (a) *Solve for* $\underline{x}$*:*

$$\begin{bmatrix} 1 & -1 & 1 & 2 \\ 2 & -2 & -1 & 3 \end{bmatrix} \underline{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

(This consists of the first two equations of No.(3).

(b) *Solve the homogeneous system for* $\underline{x}$*.*

6. (a) *What are the solutions* $\underline{x}$ *to*

$$\begin{bmatrix} 3 & 1 & 1 & 4 \\ 2 & 5 & 1 & 7 \\ 1 & 9 & 1 & 10 \end{bmatrix} \underline{x} = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

?

(b) *Solve the corresponding homogeneous system.*

7. *If you have one solution to 4a and the general solution of 4b, show that you can write down the general solution of 4a.*

8. *The same for 5a and 5b.*

9. *Solve the following system for* $\underline{x}$ *and relate your solution to that of its homogeneous counterpart.*

$$\begin{bmatrix} 2 & 1 & 3 & 5 \\ -2 & -1 & -1 & -2 \\ 4 & 2 & 7 & 11 \\ 6 & 3 & 10 & 16 \end{bmatrix} \underline{x} = \begin{bmatrix} -1 \\ 0 \\ -3 \\ -4 \end{bmatrix}$$

10. *Can you state a general theorem of which Nos.(7), (8) and (9) are special cases?*
Note that one solution to Example 2.2.3 (matrix equation (2.22)) plus the general solution (2.28) to the homogeneous Example 2.2.5 gives the general solution (2.24) to Example 2.2.3. We will return to this in the next chapter. (See Exercise 52, No.5).

11. (a) *Solve the system for* x*,* y*,* z*:*

$$\begin{bmatrix} 1 & 1 & -1 \\ -2 & 3 & 4 \\ -1 & 4 & 1 \\ 4 & -1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -3 \\ 7 \\ -2 \\ 2 \end{bmatrix}$$

(b) *Find the homogeneous solution to (a).*

12. *Reduce to row EF*
$$\begin{bmatrix} 2 & 1 & 3 & 5 \\ -2 & -1 & -1 & -2 \\ 4 & 2 & 7 & 11 \\ 6 & 3 & 10 & 16 \end{bmatrix}$$

13. ** (Allenby, p.30) For any a, b, c and d discuss the solution $\underline{x}$ of*
$$\begin{bmatrix} 1 & 2 & -1 & 3 & -1 \\ 3 & 4 & 2 & 5 & -2 \\ 1 & 0 & 4 & -1 & 0 \\ 3 & 2 & 7 & 1 & -1 \end{bmatrix} \underline{x} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$
Find also the corresponding homogeneous solution.

14. *For which scalars α and β is the system*
$$\begin{bmatrix} 1 & 2 & -3 \\ 3 & -1 & 2 \\ 1 & -5 & 8 \end{bmatrix} \underline{x} = \begin{bmatrix} \alpha \\ \beta \\ -\alpha \end{bmatrix}$$
consistent? (recall that a system of equations is consistent (or compatible) if it has at least one solution). If so, find the general solution $\underline{x}$.

15. *Show that the system*
$$\begin{bmatrix} 1 & 1 & \lambda \\ 3 & 4 & 2 \\ 2 & 3 & -1 \end{bmatrix} \underline{x} = \begin{bmatrix} 2 \\ \lambda \\ 1 \end{bmatrix}$$
is consistent for any scalar λ. Find those values of λ for which

    i the system has a unique solution $\underline{x}$ and

    ii more than one solution $\underline{x}$.

16. *Let $A\underline{x} = \underline{b}$ represent a system of m equations for n unknowns. Which of the following is true? Give reasons.*

    (a) The system always has a solution if $m < n$.

    (b) The system always has a solution if $m = n$.

    (c) The system never has a solution if $m > n$ (more equations than unknowns).

    (d) The system always has a solution if it is homogeneous.

2.2.11 Row Reduced Echelon Form (row REF)

A matrix $A = [a_{ij}]$ is said to be in **row reduced echelon form** (row REF) if

1. It is in row echelon form.
2. If a_{ij} (the pivot entry) is the first non-zero entry in row i, then
    (a) $a_{ij} = 1$ and
    (b) all other entries in column j are zeros.

2.2.12 Some examples in row reduced echelon form

$$\begin{bmatrix} 1 & -3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 3 & -5 & 0 & 0 & 7 \\ 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{bmatrix}$$

2.2.13 Any matrix can be brought into row reduced echelon form by suitable elementary row operations

Let A be in row echelon form. Suppose that A has r non-zero rows. We first divide each non-zero row by its leading (first non-zero) entry. So we may suppose that the leading entry of row r is $a_{rj} = 1$. Use row r for reference row and a_{rj} as a pivot element to reduce all entries above it to zeros. Now repeat the process with row $r-1$ etc., the last reference row being row 2. The resulting array is in row reduced echelon form (row REF).

Note that in converting a row EF to a row REF all the pivot elements remain in the same positions but become 1s.
The process is known as the **Gauss-Jordan** method.

Remark 37 *Although a matrix does not in general have a unique row echelon form (which ones do?), it can be shown that the row reduced echelon form is unique.*

2.2.14 Some examples of Gauss-Jordan reduction

Example 2.2.6 *For our first example let us reduce the array of Example 2.2.1 to row reduced echelon form. We found the row EF*

$$\begin{array}{cc} 1 & \frac{10}{3} \quad -\frac{2}{3} \\ 0 & -\frac{55}{3} \quad \frac{32}{3} \end{array}$$

Step 1:

$$\begin{array}{cc} 1 & \frac{10}{3} \quad -\frac{2}{3} \\ 0 & 1 \quad -\frac{32}{55} \end{array} \qquad \left(-\tfrac{3}{55}\right) R_2 \quad \text{ref row}$$

Step 2:

$$\begin{array}{cc} 1 & 0 \quad \frac{14}{11} \\ 0 & 1 \quad -\frac{32}{55} \end{array} \qquad -\tfrac{10}{3} R_2 + R_1$$

Example 2.2.7 *As a second example consider Example 2.1.2. Its row echelon form was found in (2.16):*

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & -2 \end{array}$$

The following steps bring it into row REF.
Step 1:

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & -2 \end{array} \qquad \begin{array}{c} \\ \frac{1}{2} R_2 \\ \text{ref row, pivot } a_{33} = 1 \end{array}$$

Step 2:

$$\begin{array}{cccc} 1 & 1 & 0 & 5 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array} \qquad \begin{array}{c} -3R_3 + R_1 \\ -\frac{1}{2} R_3 + R_2 \quad \text{new ref row, pivot } a_{22} = 1 \\ \end{array}$$

Step 3

$$\begin{array}{cccc} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array} \qquad -R_2 + R_1$$

We read off the same solution (2.3) as before, only that it is easier to do.

Example 2.2.8 *Consider Example 2.1.3 again. From (2.18) we already have the row EF of detached coefficients:*

$$\begin{array}{ccccc} 1 & 1 & 3 & 2 & -1 \\ 0 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 & -2 \end{array}$$

The following steps bring it into row REF.

Step 1

$$\begin{array}{ccccc} 1 & 1 & 3 & 2 & -1 & \\ 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2}R_2 \\ 0 & 0 & 1 & -1 & -2 & \text{ref row, pivot } a_{33}=1 \end{array}$$

Step 2

$$\begin{array}{ccccc} 1 & 1 & 0 & 5 & 5 & -3R_3+R_1 \\ 0 & 1 & 0 & 1 & 1 & -\frac{1}{2}R_3+R_2 \quad \text{new ref row, pivot } a_{22}=1 \\ 0 & 0 & 1 & -1 & -2 & \end{array}$$

Step 3

$$\begin{array}{ccccc} 1 & 0 & 0 & 4 & 4 & -R_2+R_1 \\ 0 & 1 & 0 & 1 & 1 & \\ 0 & 0 & 1 & -1 & -2 & \end{array}$$

Reading off the solution gives (unsurprisingly) the same solution (2.5) that we found before.

Example 2.2.9 *Example 2.1.4 (equation (2.7)) has a row echelon form (2.20):*

$$\begin{array}{cccc} 1 & 1 & 3 & -1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 7 \end{array}$$

The following steps bring it into row REF

$$\begin{array}{cccc} 1 & 1 & 3 & -1 & \\ 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2}R_2 \\ 0 & 0 & 0 & 1 & \frac{1}{7}R_3 \end{array}$$

$$\begin{array}{cccc} 1 & 1 & 3 & 0 & R_3+R_1 \\ 0 & 1 & \frac{1}{2} & 0 & \\ 0 & 0 & 0 & 1 & \end{array}$$

$$\begin{array}{cccc} 1 & 0 & \frac{5}{2} & 0 & -R_2+R_1 \\ 0 & 1 & \frac{1}{2} & 0 & \\ 0 & 0 & 0 & 1 & \end{array} \tag{2.29}$$

Example 2.2.10 *As a final example, consider Example 2.2.3. We found the row echelon form (2.23):*

$$\begin{array}{ccccc} 1 & -1 & 1 & 2 & 4 & \\ 0 & 1 & 3 & 1 & 8 & \\ 0 & 0 & 1 & -2 & 5 & \textit{ref row} \\ 0 & 0 & 0 & 0 & 0 & \end{array}$$

It is now two small steps to getting the reduced form:

Step 1

$$\begin{array}{ccccc} 1 & -1 & 0 & 4 & -1 & -R_3+R_1 \\ 0 & 1 & 0 & 7 & -7 & -3R_3+R_2 \quad \text{ref row} \\ 0 & 0 & 1 & -2 & 5 & \\ 0 & 0 & 0 & 0 & 0 & \end{array}$$

Step 2

$$\begin{array}{ccccc} 1 & 0 & 0 & 11 & -8 \\ 0 & 1 & 0 & 7 & -7 \\ 0 & 0 & 1 & -2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array} \qquad R_2 + R_1$$

The array of detached coefficients of Example 2.2.3 therefore has row REF

$$\begin{array}{ccccc} 1 & 0 & 0 & 11 & -8 \\ 0 & 1 & 0 & 7 & -7 \\ 0 & 0 & 1 & -2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

The required solution to Example 2.2.3 (matrix equation (2.22)) is (unsurprisingly) as before. We note that the row REF of the coefficient matrix of the same example is

$$\begin{array}{cccc} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{array} \tag{2.30}$$

The solution (2.28) to Example 2.2.5 can again be read off this array.

2.2.15 A homogeneous system with more unknowns than equations always has a non-trivial solution

More precisely, we will prove the following

Theorem: Let $A\underline{x} = \underline{0}$ be a homogeneous system of m equations in n unknowns and suppose that the row echelon form of A has $r < n$ non-zero rows.
Then there are solutions $\underline{x}$ in which the x_j corresponding to the $k = n-r$ non-pivot columns can have arbitrary values. In other words, **the solution set has k independent parameters**.

Proof
Consider the row REF A'' of A. The matrix A'' has r non-zero rows and r pivot columns. Thus there are $k = n - r$ non-pivot columns. By the nature of the row REF, the k variables x_j corresponding to these columns can be given any values, and we can then solve uniquely for the remaining r variables. (See the illustrative examples below.)

Corollary
A homogeneous system with more unknowns than equations always has a non-trivial solution.
Proof
With the above notation, we have $m < n$ and since the number r of non-zero rows in a row REF satisfies $r \leq m$, we have $r < n$.

Note 38 *In Corollary 91 of section 3.6 in chapter 3 there is an alternative approach to the topic.*

Some illustrative examples

1. Suppose that the row REF of a matrix A is

$$A'' = \begin{bmatrix} 1 & 3 & -5 & 0 & 0 & 7 \\ 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{bmatrix}$$

Here the non-pivot columns are columns 2, 3 and 6. The general solution to $A\underline{x} = \underline{0}$ is

$$\underline{x} = \begin{bmatrix} -3x_2 + 5x_3 - 7x_6 \\ x_2 \\ x_3 \\ 2x_6 \\ -3x_6 \\ x_6 \end{bmatrix} = \begin{bmatrix} -3t_1 + 5t_2 - 7t_3 \\ t_1 \\ t_2 \\ 2t_3 \\ -3t_3 \\ t_3 \end{bmatrix}$$

We have renamed the parameters as $t_1 = x_2$, $t_2 = x_3$ and $t_3 = x_6$.

2. As a second example consider Example 2.2.5 (equation 2.27) which in REF reads (see array (2.30)):

$$\begin{bmatrix} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \underline{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This shows that we are actually dealing with 3 equations for 4 unknowns. Column 4 of the coefficient matrix is the only non-pivot column. Hence $x_4 = t$e can have any value and we get the same solution as before (equation (2.28).

Exercise 39 *1. Prove that if the augmented array $[A, \underline{b}]$ for a system of equations $A\underline{x} = \underline{b}$ has been reduced to row (reduced) echelon form $[A', \underline{b}']$ then A' is also in (reduced) echelon form.*

Hint 40 *A zero row of A' comes after a non-zero row of A'.*

2. Find the row reduced echelon form of the matrices and arrays of detached coefficients in Exercise (35), numbers 1a, 1b, 2, 3, 4, 9, 11 , 12 and 13. Solve (once more) the problems involving systems of equations $A\underline{x} = \underline{b}$ using the row REF. Write down the row REF of the corresponding coefficient matrix A. Illustrate the above theorem 2.2.15 in the homogeneous cases.

3. Consider the general problem of solving two linear equations for two unknowns x_1 and x_2:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{2.31}$$

(a) In (2.31) all the entries except x_1 and x_2 are supposed known. Using formal Gauss-reduction find formulae for the unknowns x_1 and x_2 in terms of the other symbols. For the purposes of reduction you may assume any number you like is non-zero. In fact, conclude that if only $d = a_{11}a_{22} - a_{21}a_{12} \neq 0$, then you have a formula for $\underline{x}$. Can you see that the solution must be unique?

Remark 41 *The number $a_{11}a_{22} - a_{21}a_{12}$ is called the* ***determinant*** *of the matrix $A = [a_{ij}]$. and is denoted by $|A|$. For example,*

$$\begin{vmatrix} 3 & -2 \\ 7 & 4 \end{vmatrix} = (3)(4) - (7)(-2) = 26$$

We will study determinants in some detail in Chapter 4.

Hint 42 *Consider the array of detached coefficients and, assuming $a_{11} \neq 0$, the step*

$$\begin{matrix} a_{11} & a_{12} & b_1 & \\ 0 & -\frac{a_{21}a_{12}}{a_{11}} + a_{22} & -\frac{a_{21}b_1}{a_{11}} + b_2 & -\frac{a_{21}}{a_{11}}R_1 + R_2 \end{matrix} \tag{2.32}$$

Now assume

$$-\frac{a_{21}a_{12}}{a_{11}} + a_{22} \neq 0$$

and solve for x_2. Next, find x_1.

(b) Use your formula to solve for $\underline{x}$:

$$\begin{bmatrix} -5 & 11 \\ 3 & -7 \end{bmatrix} \underline{x} = \begin{bmatrix} -2 \\ 13 \end{bmatrix}$$

2.3 Summary of Chapter 2

2.3.1 Matrix equation of simultaneous equations

Given a system of m equations for n unknowns, this is expressed compactly by a matrix equation (See section 2.2).

$$A\underline{x} = \underline{b}$$

where A is the $m \times n$ coefficient matrix, $\underline{b}$ is a column with m entries and $\underline{x}$ is a column with n entries.

Entry i of the column matrix $A\underline{x}$ is the 'dot' product of row i of A and the column $\underline{x}$ (See subsection 2.2.3 and equation 2.13):

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{i1}x_1 & + & \cdots & + & a_{in}x_n \end{bmatrix}$$

2.3.2 Detached Coefficients

(See subsection 2.2.5).
The other way to express the system of equations $A\underline{x} = \underline{b}$ is as the augmented matrix, or array of detached coefficients

$$C = [A, \underline{b}]$$

where the last column of R is understood to be the column $\underline{b}$ of constants.

2.3.3 Elementary row operations (eros). Row echelon and row reduced echelon form

(See subsections 2.2.6, 2.2.8 and 2.2.11).

Elementary row operations (eros), known as Gauss reduction, are done on the array R reducing it to row echelon form (row EF):

$$C' = \left[A', \underline{b}'\right]$$

Further eros (Gauss-Jordan) done on C' reduce it to row reduced echelon form (row REF):

$$C'' = \left[A'', \underline{b}''\right]$$

The important fact about elementary row operations on a system of equations (or its detached array of coefficients) is that they leave **the solution set unchanged**. For any particular $\underline{x}$ the equation $A\underline{x} = \underline{b}$ is satisfied if, and only if, the equation $A'\underline{x} = \underline{b}'$ is satisfied if, and only if, the equation $A''\underline{x} = \underline{b}''$ is satisfied.

2.3.4 Homogeneous equations

(See subsection 2.2.10).

A homogeneous system of equations has $\underline{b} = \underline{0}$ and consequently $\underline{b}' = \underline{b}'' = \underline{0}$, since an elementary row operation done on $\underline{0}$ leaves it as $\underline{0}$.

For a particular $\underline{x}$, the matrix equation $A\underline{x} = \underline{0}$ holds if, and only if, $A'\underline{x} = \underline{0}$ if, and only if, $A''\underline{x} = \underline{0}$ holds.

A system of homogeneous equations with more unknowns than equations always has a non-trivial solution (see 2.2.15).

Chapter 3

Linear Transformations and Matrices

In this chapter we will learn how the solution to a set of linear equations is the **inverse operation** of the interpretation of a matrix A as a **transformation** or **mapping**. In this sense A transforms a vector $\underline{x}$ into another vector $\underline{y} = A\underline{x}$. Examples in ordinary space are projections on planes and lines through the origin, reflections in such objects and rotations about an axis through O (important for mechanics).
Next, we deal with the difficult idea of linear independence, and it is followed by a section on subspaces that can be regarded as an introduction to the idea of a general 'vector space'. Finally, there is a section on inverses of matrices.

3.1 The Algebra of Vectors and Matrices

We begin by developing some systematic notation for $m \times n$ matrices, their rows and columns and for column and row vectors in general.

3.1.1 Systematic Notation For Matrices

1. Recall from section 2.2 that an $m \times n$ **matrix** is a rectangular A with m rows and n columns. In standard notation, $A = [a_{ij}]$ where the $i-j$ entry $a_{ij} \in \Re$ is at the intersection of row i and column j.
   The **set** of $m \times n$ matrices is denoted by $\Re^{m\times n}$. Hence $A \in \Re^{m\times n}$ means that A has m rows and n columns and that the **size** of A is $m \times n$. Some authors write $A = A_{m\times n}$ to indicate the size of A.

   Alternative notation If $A = [a_{ij}]$, we will also find it most convenient to use the notation

   $$[A]_{ij} = a_{ij} \tag{3.1}$$

   to describe the $i-j$ entry a_{ij} of A.

   An $n \times n$ matrix is called **square** of size n. Such matrices deserve to be studied in their own right (see section 3.5).

2. From the definitions it follows that $\Re^{n\times 1}$ is the set of columns with n entries. We call

these **column vectors** and write

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{3.2}$$

$\Re^{1\times n}$ denotes the set of **row vectors** with n entries:

$$\underline{a} = \begin{bmatrix} a_1 & a_2 & ... & a_n \end{bmatrix} \tag{3.3}$$

For typographical reasons this row vector is sometimes written $[a_1, ..., a_n]$. It is traditional to use $\Re^n$ for both $\Re^{n\times 1}$ (column vectors) and $\Re^{1\times n}$ (row vectors). We will only use the notation $\Re^n$ if the context makes it quite clear whether rows or columns are meant.

3. The $m \times n$ **zero matrix** has 0 for all its entries and is written $O_{m\times n}$, or simply as O when the size is understood.

4. **Addition of matrices**
Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be of the same $m \times n$ size. Their sum $A + B = C$ where the (i, j)-entry of C is $c_{ij} = a_{ij} + b_{ij}$. In brief, using the notation or equation (3.1),

$$[A + B]_{ij} = [A]_{ij} + [B]_{ij} \tag{3.4}$$

It is essential to note that the addition $A+B$ makes sense only if A and B have the same size.

5. **Multiplication of a matrix by a scalar**.
Let $A = [a_{ij}]$ and $s \in \Re$. The product $sA = As$ has the number sa_{ij} in the (i, j) −position:

$$[sA]_{ij} = [As]_{ij} = s\,[A]_{ij} = [A]_{ij}\,s$$

Remark 43 *Notice that it makes no difference if we write the scalar s on the left or on the right of A. The matrix $-A$ is defined as $(-1)\,A$ and has for (i, j) entry the number $-a_{ij}$ i.e., $[(-1)\,A]_{ij} = -[A]_{ij}$. The difference $A - B$ of two matrices of the same size means the same thing as $A + (-B)$.*

*We will see that if $\underline{a}$ is a **row** vector (3.3) then it is more natural to write $s\underline{a} = \begin{bmatrix} sa_1 & sa_2 & ... & sa_n \end{bmatrix}$ rather than $\underline{a}s$.*
*On the other hand, for a **column** $\underline{b}$, as in equation (3.2), it is better to write $\underline{b}s$. See, for example, Equation (3.23) below.*

6. **Row** i of the $m \times n$ matrix $A = [a_{ij}]$ is denoted by $A_{i\bullet}$. Hence,

$$A_{i\bullet} = \begin{bmatrix} a_{i1} & a_{i2} & .... & a_{in} \end{bmatrix} \quad (i = 1, \cdots, m) \tag{3.5}$$

7. **Column** j of the $m \times n$ matrix $A = [a_{ij}]$ is denoted by $A_{\bullet j}$. Hence,

$$A_{\bullet j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad (j = 1, \cdots, n) \tag{3.6}$$

Remark 44 *The symbols for row i and column j are consistent with our other notation. Thus $[A]_{ij} = a_{ij}$ is at the **intersection** of row $A_{i\bullet}$ and column $A_{\bullet j}$.*

Other notations in use for the i^{th} row of A, are A_i and $[A]_i$, and for the j^{th} column of A, $A^{(j)}$ or $[A]^j$. However, we will not use these.

8. The **unit column vectors** $\underline{e}_j \in \Re^{n\times 1}$ $(j = 1, ..., n)$ have a 1 in position j and zeros elsewhere:

$$\underline{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \text{row } j \tag{3.7}$$

9. **The** $n \times n$ **Identity matrix** is the $n \times n$ matrix with columns $\underline{e}_1$, ... ,$\underline{e}_n$ (in that order). It is denoted by I_n (or simply I if n is understood):

$$I_n = \begin{bmatrix} 1 & 0 & \vdots & 0 \\ 0 & 1 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & 1 \end{bmatrix} \tag{3.8}$$

As we will see, I_n behaves like the number 1 when we multiply matrices. (See Exercise 63 No.5). We note that, writing $I = I_n$, by definition $I_{\bullet j} = \underline{e}_j$ for $j = 1, \cdots, n$. The **unit row vectors** are

$$I_{i\bullet} = \begin{bmatrix} 0 & \cdots & 1 & (\text{position } i) & \cdots & 0 \end{bmatrix} \qquad (i = 1, \cdots, n)$$

The identity $I = I_n$ of (3.8) is equivalently defined by

$$[I]_{ij} = 1 \text{ if } i = j \text{ and } [I]_{ij} = 0 \text{ if } i \neq j \qquad (i, j = 1, \cdots, n)$$

10. The **transpose** of the $m \times n$ matrix $A = [a_{ij}]$ is denoted by A^T ("A transpose"). The transpose A^T has for its (i, j) entry the number a_{ji}. Thus A^T is an $n \times m$ matrix and can also be vaguely defined as 'the matrix A with its rows and columns interchanged'. In fact, using the notation (3.1),

$$\left[A^T\right]_{ji} = [A]_{ij} \quad (i = 1, ..., n;\ j = 1, 2, ..., m) \tag{3.9}$$

11. The **product** $\underline{a}\,\underline{b}$ of the **row** $\underline{a}$ with n entries and the **column** $\underline{b}$ with n of entries is

$$\underline{a}\,\underline{b} = \begin{bmatrix} a_1 & a_2 & ... & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \tag{3.10}$$

Remark 45 *This was introduced in subsection 2.2.3 of Chapter 2. You can, if you wish, call this a 'dot product' since it is a generalization of the dot product Equation (1.4) of Chapter 1, but there is no need here for the 'dot'. Note the order in* $\underline{a}\,\underline{b}$*: first* $\underline{a}$*, then* $\underline{b}$*, This is important as* $\underline{b}\underline{a}$ *will turn out to be something quite different: in fact, an* $n \times n$ *matrix.(See No.3 of Exercise 63).*

3.1.2 Linear combinations of vectors, linear dependency, Span

1. The column vector $\underline{v} \in \Re^{n\times 1}$ is a **linear combination** of the vectors $\underline{u}_1$, $\underline{u}_2$,...,$\underline{u}_k$ with **coefficients** s_1, s_2, ..., s_k if

$$\underline{v} = \underline{u}_1 s_1 + \underline{u}_2 s_2 + \cdots + \underline{u}_k s_k \tag{3.11}$$

We then say that $\underline{v}$ **depends linearly** on $\underline{u}_1$, $\underline{u}_2$,...,$\underline{u}_k$. Of course, we can also write $\underline{v} = s_1\underline{u}_1 + s_2\underline{u}_2 + \cdots + s_k\underline{u}_k$, with the coefficients on the left, but, as remarked, we prefer to write the coefficients of columns on the right. If $\underline{v}$ and $\underline{u}_1$, $\underline{u}_2$,...,$\underline{u}_k$ are row vectors all of the same size our preference would be

$$\underline{v} = s_1\underline{u}_1 + s_2\underline{u}_2 + \cdots + s_k\underline{u}_k$$

2. The idea of 'span' was introduced in Exercise 2, No.14.
The span of (column or row) vectors $\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k$ is the set of all linear combinations of these vectors. It is also called the **space spanned by** the vectors and is denoted by $sp(\underline{u}_1, \ldots, u_k)$. Thus,

$$\underline{v} \in sp(\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k)$$

if, and only if, (3.11) holds for certain coefficients $s_1, \ldots, s_k$.

The idea of 'span' first came up in No.14 of Exercise 2.

The **column space** of a matrix A is the span of its columns; its **row space** is the space spanned by its rows.

The ideas of 'span', column space and row space will only be systematically developed in the section 3.6 on vector spaces.

3. **Linear combinations of matrices**

We wish to extend the definition of 'linear combination' to matrices: Let A_1, A_2, ..A_k be k matrices, all of the same size and suppose that s_1, s_2, ..., s_k are k scalars. The linear combination of A_1, A_2, ..,A_k with coefficients s_1, s_2, ..., s_k is

$$s_1A_1 + s_2A_2 + ... + s_kA_k$$

3.1.3 Some examples

1. $\begin{bmatrix} -1 & \frac{5}{4} & 13 \\ 2 & -4 & -7 \end{bmatrix} \in \Re^{2\times 3}$, $\begin{bmatrix} -4 & 5 \\ 5 & \frac{2}{3} \\ 11 & \frac{7}{2} \end{bmatrix} \in \Re^{3\times 2}$, $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \in \Re^{2\times 2}$

2. $\begin{bmatrix} -\frac{3}{7} \\ \frac{2}{5} \end{bmatrix} \in \Re^{2\times 1}$, $\begin{bmatrix} \frac{5}{3} \\ -\frac{2}{9} \\ 6 \end{bmatrix} \in \Re^{3\times 1}$

3. $$O_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; O_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

4. $$\begin{bmatrix} -1 & \frac{5}{4} & 13 \\ 2 & -4 & -7 \end{bmatrix} + \begin{bmatrix} 2 & -\frac{7}{4} & -7 \\ -3 & \frac{5}{2} & -7 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} & 6 \\ -1 & -\frac{3}{2} & -14 \end{bmatrix}$$

5. $$\begin{bmatrix} -1 & \frac{5}{4} & 13 \\ 2 & -4 & -7 \end{bmatrix} (100) = (100) \begin{bmatrix} -1 & \frac{5}{4} & 13 \\ 2 & -4 & -7 \end{bmatrix} = \begin{bmatrix} -100 & 125 & 1300 \\ 200 & -400 & -700 \end{bmatrix}$$

6. (a) Let

$$A = \begin{bmatrix} -1 & 2 & -\frac{2}{3} \\ \frac{5}{4} & -4 & 5 \\ 13 & -7 & \frac{3}{8} \end{bmatrix} \tag{3.12}$$

Then

$$A_{1\bullet} = \begin{bmatrix} -1 & 2 & -\frac{2}{3} \end{bmatrix}$$
$$A_{3\bullet} = \begin{bmatrix} 13 & -7 & \frac{3}{8} \end{bmatrix}$$

(b) Let

$$X = \begin{bmatrix} \frac{7}{2} \\ -\frac{3}{2} \\ 9 \\ -\frac{5}{6} \end{bmatrix}$$

then

$$X_{2\bullet} = \begin{bmatrix} -\frac{3}{2} \end{bmatrix}$$

$$X_{4\bullet} = \begin{bmatrix} -\frac{5}{6} \end{bmatrix}$$

Remark 46 *We usually use an 'underline' notation like (3.2) for column vectors but will occasionally write $X_2 = X_{2\bullet} = -\frac{3}{2}$ if it is understood that X is a column matrix (i.e. a column vector). Similar remarks apply to row vectors (i.e. row matrices).*

7. (a) If A is equation (3.12) above then

$$A_{\bullet 2} = \begin{bmatrix} 2 \\ -4 \\ -7 \end{bmatrix}$$

$$A_{\bullet 1} = \begin{bmatrix} -1 \\ \frac{5}{4} \\ 13 \end{bmatrix}$$

(b) If

$$Y = \begin{bmatrix} 5 & -\frac{7}{2} & -9 & 10 & -23 \end{bmatrix}$$

$$Y_{\bullet 2} = \begin{bmatrix} -\frac{7}{2} \end{bmatrix}$$

$$Y_{\bullet 5} = [-23]$$

8. (a) The vectors (3×1 column matrices) $\underline{e}_1$, $\underline{e}_2$ and $\underline{e}_3$ in $\Re^{3\times 1}$ are

$$\underline{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \underline{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \ \underline{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

We met these as **rows** in Chapter 1, 1.1.17, calling them $\underline{i}$, $\underline{j}$, and $\underline{k}$ respectively.

(b) In $\Re^{4\times 1}$

$$\underline{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \underline{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ \underline{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ \underline{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

9. We have

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

10. (a) Let

$$A = \begin{bmatrix} -5 & 0 & 2 \\ 7 & 6 & -1 \\ \frac{1}{2} & -1 & 3 \\ -\frac{1}{13} & 9 & 0 \end{bmatrix}$$

then $A \in \Re^{4\times 3}$, while the transpose $A^T \in \Re^{3\times 4}$ and

$$A^T = \begin{bmatrix} -5 & 7 & \frac{1}{2} & -\frac{1}{13} \\ 0 & 6 & -1 & 9 \\ 2 & -1 & 3 & 0 \end{bmatrix}$$

(b) The transpose of the column vector (3.2) is the row

$$\underline{b}^T = \begin{bmatrix} b_1 & b_2 & \ldots & b_n \end{bmatrix}$$

and the transpose of the row vector (3.5) of A is a column vector.

(c) $[I_{i\bullet}]^T = I_{\bullet i}$ and $[I_{\bullet i}]^T = I_{i\bullet}$ and $I^T = I$, where $I = I_n$.

11. (a) An example of the product $\underline{a}\,\underline{b}$ of a row $\underline{a}$ and a column $\underline{b}$ is

$$\begin{bmatrix} -3 & 5 & -7 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ -2 \\ 9 \end{bmatrix} = (-3)(2) + (5)(3) + (-7)(-2) + (4)(9) = 59$$

(b) Another, using a general 2×3 matrix $A = [a_{ij}]$, and a general 3×4 matrix $B = [b_{ij}]$ is

$$A_{i\bullet} B_{\bullet j} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} \qquad (i = 1, 2;\ j = 1, 2, 3, 4)$$

Note that such a product is only possible when the number of columns of A equals the number of rows of B. The significance of such products will be seen in subsection 3.3.4 below.

Referring to linear combinations (item (3) in 3.1.1) we have:

1. (a) In $\Re^{4\times 1}$

$$\begin{bmatrix} -5 \\ \frac{1}{2} \\ -3 \\ \frac{2}{3} \end{bmatrix} (-6) + \begin{bmatrix} \frac{7}{2} \\ -\frac{3}{2} \\ 9 \\ -\frac{5}{6} \end{bmatrix} (3) + \begin{bmatrix} 11 \\ 2 \\ -4 \\ -1 \end{bmatrix} \left(-\frac{1}{2}\right) = \begin{bmatrix} 35 \\ -\frac{17}{2} \\ 47 \\ -6 \end{bmatrix}$$

(b) In $\Re^{1\times 5}$

$$\begin{aligned} & (-2)\begin{bmatrix} p & q & \frac{1}{2} & -3 & s \end{bmatrix} + 4\begin{bmatrix} -1 & r+2 & -\frac{1}{4} & \frac{5}{2} & 1 \end{bmatrix} \\ = \ & \begin{bmatrix} -2p-4 & -2q+4r+8 & -2 & 16 & -2s+4 \end{bmatrix} \end{aligned}$$

2. The span of the rows of I_3 is clearly $\Re^{1\times 3}$.
Consider the column space of matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The span of the columns of A is the whole of $\Re^{4\times 1}$, since for any choices of b_1, b_2, b_3 and b_4,

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} (b_1 - b_2) + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} (b_2 - b_3) + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} (b_3 - b_4) + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} b_4$$

See also subsections 3.1.6 and 3.1.9 below.

3. In $\Re^{2\times 2}$

$$\begin{aligned} (7)\begin{bmatrix} -1 & 2 \\ 3 & 6 \end{bmatrix} + (-3)\begin{bmatrix} 0 & 8 \\ -2 & 5 \end{bmatrix} &= \begin{bmatrix} -1 & 2 \\ 3 & 6 \end{bmatrix}(7) + \begin{bmatrix} 0 & 8 \\ -2 & 5 \end{bmatrix}(-3) \\ &= \begin{bmatrix} -7 & -10 \\ 27 & 27 \end{bmatrix} \end{aligned}$$

3.1.4 Some basic algebraic properties

For all $m \times n$ matrices A, B and C and scalars s and t,

1. $A + (B + C) = (A + B) + C$ (associative law for addition).
2. $A + B = B + A$ (commutative law for addition).
3. $O + A = A$ ($O = O_{m\times n}$ behaves *like* the number 0).
4. $A + (-A) = O$.
5. $sA = As$ (by definition).
6. $s(A + B) = sA + sB$.
7. $(s + t)A = sA + tA$.
8. $s(tA) = (st)A$.
9. $1A = A$.
10. Every vector $\underline{x} \in \Re^{n\times 1}$ can be uniquely written as a linear combination

$$\underline{x} = \underline{e}_1 x_1 + \underline{e}_2 x_2 + ... + \underline{e}_n x_n$$

(Compare 1.1.17 in Chapter 1). Hence it is trivial that every $\underline{x}$ depends linearly on the unit vectors $\underline{e}_1$, $\underline{e}_2$,..., $\underline{e}_n$, in other words that these unit vectors span $\Re^{n\times 1}$. However, there are many other sets of vectors that have a similar property, as we shall see. In Chapter 1, No.15 of Exercise 2 we indicated that any three mutually orthogonal vectors in $\Re^3$ span the whole of space. You are asked to show this rigorously in Exercise 78, No.9.

11. $\left(A^T\right)^T = A$ for any matrix A.
12. Referring to the product (3.10),

$$(\underline{a}\,\underline{b})^T = \underline{b}^T\,\underline{a}^T \tag{3.13}$$

13. (a) Let $\underline{u}$ and $\underline{v}$ be column vectors with n entries and $\underline{a}$ a row with n entries. Then

$$\underline{a}\,(\underline{u} + \underline{v}) = \underline{a}\,\underline{u} + \underline{a}\,\underline{v} \tag{3.14}$$

(b) If s is a scalar,

$$\underline{a}\,(\underline{u}s) = (\underline{a}\,\underline{u})\,s = (s\underline{a})\,\underline{u} \tag{3.15}$$

(c) Equations (3.14) and (3.15) combine so that for all scalars s and t:

$$\underline{a}\ (\underline{u}s + \underline{v}t) = (\underline{a}\,\underline{u})\,s + (\underline{a}\,\underline{v})\,t \tag{3.16}$$

(d) More generally, Equation (3.16) can be generalized as follows: Suppose $\underline{u}_1$, $\underline{u}_2$, ..,$\underline{u}_k$ are column vectors in $\Re^{n\times 1}$ and s_1, s_2, ..,.s_k are scalars and $\underline{a} \in \Re^{1\times n}$ is a row vector. Then

$$\underline{a}\ (\underline{u}_1 s_1 + \underline{u}_2 s_2 + ... + \underline{u}_k s_k) = (\underline{a}\,\underline{u}_1)\,s_1 + (\underline{a}\,\underline{u}_2)\,s_2 + ... + (\underline{a}\,\underline{u}_k)\,s_k \tag{3.17}$$

14. (a) Quite similarly, if $\underline{a}$ and $\underline{b}$ are row vectors with n entries and $\underline{u}$ is a column with n entries,

$$(\underline{a} + \underline{b})\ \underline{u} = \underline{a}\,\underline{u} + \underline{b}\,\underline{u} \tag{3.18}$$

(b) and for scalars t,

$$(t\underline{a})\ \underline{u} = t\,(\underline{a}\,\underline{u}) \tag{3.19}$$

(c) Equations (3.18) and (3.19) combine so that for all scalars s and t:

$$(s\underline{a} + t\underline{b})\ \underline{u} = s\,(\underline{a}\,\underline{u}) + t\,(\underline{b}\,\underline{u}) \tag{3.20}$$

(d) The result (3.20) can be generalized as follows: Suppose $\underline{a}_1$, $\underline{a}_2$, ..,$\underline{a}_k$ are rows in $\Re^{1\times n}$ and t_1, t_2, ..,.t_k are scalars and $\underline{u} \in \Re^{n\times 1}$ is a column, then

$$(t_1\underline{a}_1 + t_2\underline{a}_2 + ... + t_k\underline{a}_k)\ \underline{u} = t_1\,(\underline{a}_1\,\underline{u}) + t_2\,(\underline{a}_2\,\underline{u}) + ... + t_k\,(\underline{a}_k\,\underline{u}) \tag{3.21}$$

3.1.5 Some proofs

The above properties and their proofs are similar to those for vectors discussed in Chapter 1, subsections 1.1.6 and 1.1.9, equation (1.8). See No.7 of Exercise 47 below.
We prove a few properties as illustrations.

Proof of property (6) that $s\,(A + B) = sA + sB$.

The (i, j) entry in $s\,(A + B)$ is $s\,(a_{ij} + b_{ij})$. But $s\,(a_{ij} + b_{ij}) = sa_{ij} + sb_{ij}$, which is the $(i, j)-$ entry in $sA + sB$. Since $s\,(A + B)$ and $sA + sB$ have the same entries, they are equal.

Using the alternate notation (equation 3.1,

$$\begin{aligned} [s\,(A + B)]_{ij} &= s\,[A + B]_{ij} = s\left([A]_{ij} + [B]_{ij}\right) \\ &= s\,[A]_{ij} + s\,[B]_{ij} = [sA]_{ij} + [sB]_{ij} = [sA + sB]_{ij} \end{aligned}$$

Proof of property (11)

$$\left[\left(A^T\right)^T\right]_{ij} = \left[A^T\right]_{ji} = [A]_{ij}$$

Proof of equation (3.14) With obvious notation, entry i in the column $\underline{u} + \underline{v}$ is $u_i + v_i$. Hence,

$$\begin{aligned} \underline{a}\ (\underline{u} + \underline{v}) &= a_1\,(u_1 + v_1) + \cdots + a_n\,(u_n + v_n) \\ &= (a_1u_1 + a_1v_1) + \cdots + (a_nu_n + a_nv_n) \\ &= (a_1u_1 + \cdots + a_nu_n) + (a_1v_1 + \cdots + a_nv_n) \\ &= \underline{a}\,\underline{u} + \underline{a}\,\underline{v} \end{aligned}$$

Or, using $\sum$- notation,

$$\begin{aligned}\underline{a}\,(\underline{u}+\underline{v}) &= \sum_{i=1}^{n} a_i\,(u_i+v_i) = \sum_{i=1}^{n}(a_iu_i+a_iv_i)\\ &= \sum_{i=1}^{n} a_iu_i + \sum_{i=1}^{n} a_iv_i\\ &= \underline{a}\,\underline{u}+\underline{a}\,\underline{v}\end{aligned}$$

Proof of equation (3.15). Again with obvious notation, the j^{th} entry in the vector $\underline{u}s$ is u_js. Hence

$$\begin{aligned}\underline{a}\,(\underline{u}s) &= a_1\,(u_1s)+\cdots+a_n\,(u_ns)\\ &= (a_1u_1+\cdots+a_nu_n)\,s\\ &= (\underline{a}\,\underline{u})\,s\end{aligned}$$

Proof of equation (3.16). By equation (3.14), $\underline{a}\,(\underline{u}s+\underline{v}t) = \underline{a}\,(\underline{u}s)+\underline{a}\,(\underline{v}t)$. Using equation (3.15), $\underline{a}\,(\underline{u}s)+\underline{a}\,(\underline{v}t) = (\underline{a}\,\underline{u})\,s+(\underline{a}\,\underline{v})\,t$.

Exercise 47 *1. Consider 23 in Chapter 2 and the matrices in 2.2.1 referred to there. For each of these as well as for the following $A\in\Re^{m\times n}$ find m and n and, using our notation in equation (3.5) and equation (3.6), write down all rows and columns of A.*

(a) $A=\begin{bmatrix}-2 & 5\\ 3 & -\frac{4}{7}\\ -1 & \frac{1}{2}\end{bmatrix}$

(b) $A=\begin{bmatrix}4 & 7 & 0 & \frac{2}{5}\\ 0 & 9 & -1 & \frac{3}{4}\\ -\frac{5}{3} & 1 & \frac{2}{3} & 0\end{bmatrix}$

(c) $A=\begin{bmatrix}p & f & c\\ b & q & p\\ d & f & r\end{bmatrix}$

(d) $A=\begin{bmatrix}-3 & -1 & 8 & 0\\ 2 & -7 & 9 & \frac{1}{2}\\ 0 & \frac{2}{3} & -2 & 0\\ 1 & -4 & 5 & -6\end{bmatrix}$

2. Write down C if it is given that

(a) C has three columns $C_{\bullet 1}=\begin{bmatrix}4\\ -5\end{bmatrix}$, $C_{\bullet 2}=\begin{bmatrix}2\\ 7\end{bmatrix}$, $C_{\bullet 3}=\begin{bmatrix}0\\ \frac{2}{3}\end{bmatrix}$

(b) C has three rows and $C_{1\bullet}=\begin{bmatrix}q & 5 & -z\end{bmatrix}$, $C_{2\bullet}=\begin{bmatrix}0 & a & b\end{bmatrix}$, $C_{3\bullet}=\begin{bmatrix}-11 & x & \frac{2}{3}\end{bmatrix}$.

(c) $C=kI_5$ (k a scalar)

3. Find:

(a) $\begin{bmatrix}-1\\ \frac{3}{2}\\ 2\\ 0\end{bmatrix}2+\begin{bmatrix}\frac{1}{2}\\ 1\\ -\frac{2}{3}\\ -\frac{1}{6}\end{bmatrix}(-6)+\begin{bmatrix}1\\ -\frac{2}{5}\\ \frac{1}{10}\\ -1\end{bmatrix}5+\begin{bmatrix}0\\ 3\\ -1\\ 4\end{bmatrix}(-1)$

(b) $4\begin{bmatrix}6 & -2\\ -1 & 8\\ 7 & 2\end{bmatrix}+3\begin{bmatrix}11 & 0\\ 2 & 1\\ -4 & 5\end{bmatrix}-2\begin{bmatrix}0 & 6\\ 9 & 4\\ -3 & 7\end{bmatrix}$

(c) $\left(\frac{1}{2}+p\right)\begin{bmatrix} 2 & -4 & 6 & 3 \end{bmatrix}-\frac{2}{3}(2p+q)\begin{bmatrix} -3 & 9 & 18 & -6 \end{bmatrix}+\frac{1}{4}(p-q)\begin{bmatrix} 0 & -24 & 8 & 6 \end{bmatrix}$

4. *Verify equation (3.16) in case* $\underline{a}=\begin{bmatrix} -3 & \frac{5}{2} & 2 & -1 \end{bmatrix}$, $\underline{u}=\begin{bmatrix} \frac{1}{2} & -1 & -3 & \frac{1}{4} \end{bmatrix}^T$, $\underline{v}=\begin{bmatrix} -\frac{5}{6} & \frac{1}{3} & -\frac{1}{2} & 1 \end{bmatrix}^T$, $s=2$ *and* $t=-6$.

5. *Find the transposes of the matrices* $\begin{bmatrix} -\frac{5}{6} & -2 & \frac{1}{2} \\ 1 & 8 & -9 \\ -7 & 2 & \frac{1}{3} \end{bmatrix}$, $\begin{bmatrix} x & -u & c & d \end{bmatrix}$ *and* $\begin{bmatrix} v & -3 & 7 & a \\ -9 & w & 0 & z \end{bmatrix}$.

6. *For the matrices A in No.5 compare* $A_{i\bullet}$ *and* $\left[A^T\right]_{\bullet i}$ *as well as* $A_{\bullet j}$ *and* $\left[A^T\right]_{j\bullet}$. *Can we say that* $[A_{i\bullet}]^T=\left[A^T\right]_{\bullet i}$ *and* $[A_{\bullet j}]^T=\left[A^T\right]_{j\bullet}$? *Are these true for any matrix A?*

7. *Prove the above basic properties in subsection 3.1.4. For equation (3.17) and equation (3.21) you may assume k=3, since the proofs of the general cases are quite similar.*

   Hint 48 *Study the above proofs in 3.1.5 as well as your own proofs of the statements in 1.1.6 and 1.1.9 of Chapter 1. Apart from the fact that Chapter 1 concentrated on vectors in* $\Re^{1\times 3}$ *or* $\Re^{1\times 2}$ *most of the proofs for general vectors go through almost word-for-word.*

8. *(a) i. In* $\Re^m=\Re^{m\times 1}$ *let* $\underline{x}$ *be a linear combination of the vectors* $\underline{a}_1$, $\underline{a}_2$ *and* $\underline{a}_3$. *Suppose that each of* $\underline{a}_1$, $\underline{a}_2$ *and* $\underline{a}_3$ *is a linear combination of* $\underline{e}$, $\underline{f}$ *and* $\underline{g}$. *Show that* $\underline{x}$ *is a linear combination of* $\underline{e}$, $\underline{f}$ *and* $\underline{g}$.

   ii. Express the this result in terms of the 'span' concept.

   (b) Is the above result 8(a)i true for $p\times q$ *matrices X,* A_1 A_2, A_3, E, F *and* G *(in place of the vectors* $\underline{x}$,...,$\underline{g}$*)? Can you prove your answer?*

   (c) Find a ***statement*** *that generalizes No.8a: Suppose that* $\underline{x}\in\Re^m$ *depends linearly on* $\underline{a}_1,\underline{a}_2,\ldots,\underline{a}_k$ *and each vector* $\underline{a}_j$ *depends linearly on* $\underline{b}_1,\underline{b}_2,\ldots,\underline{b}_n$, *then ...? (You need not prove your statement yet, but see Exercise 63, No.13).*

9. *Show that, provided A and B are of the same size,* $(A+B)^T=A^T+B^T$. *(Use the notation of equation 3.1).*

   Solution*:*
   $[(A+B)^T]_{ji}=[A+B]_{ij}=[A]_{ij}+[B]_{ij}=[A^T]_{ji}+[B^T]_{ji}=[A^T+B^T]_{ji}$.

10. *For the following pairs A, B of matrices find all possible products* $A_{i\bullet}B_{\bullet j}$ *and arrange them in matrix form in what you think is a nice way.*

    (a)
    $$A=\begin{bmatrix} 2 & -9 \\ 3 & 1 \end{bmatrix},\quad B=\begin{bmatrix} 8 & 2 & -1 \\ -5 & 4 & 7 \end{bmatrix}$$

    (b)
    $$A=\begin{bmatrix} -1 & 5 \\ 3 & 0 \\ 2 & -4 \end{bmatrix},\quad B=\begin{bmatrix} -1 & 2 & 1 & 0 \\ 3 & 0 & -4 & 7 \end{bmatrix}$$

    Solution*:*
    $\begin{bmatrix} 61 & -32 & -65 \\ 19 & 10 & 4 \end{bmatrix}$

    $$A=\begin{bmatrix} -1 & 5 \\ 3 & 0 \\ 2 & -4 \end{bmatrix},\quad B=\begin{bmatrix} -1 & 2 & 1 & 0 \\ 3 & 0 & -4 & 7 \end{bmatrix}$$

11. Show that with appropriate definitions in $\Re^{n\times 1}$ (or in $\Re^{1\times n}$) properties 1 - 5 of subsection 1.1.13 in Chapter 1 are valid. Notice in particular No.4, which allows one to define the length $|\underline{x}|$ of a vector $\underline{x} \in \Re^{n\times 1}$ as $|\underline{x}| = \sqrt{\underline{x}^T\underline{x}}$. This in turn gives us a definition of the distance between $\underline{a}$ and $\underline{b}$ as $|\underline{a} - \underline{b}|$.

3.1.6 $A\underline{x}$ viewed in two equivalent ways

1. In terms of the products $A_{i\bullet}\underline{x}$:

   Let $A = [a_{ij}]$ be $m \times n$ and $\underline{x} = [x_j]$ be $n \times 1$. In Chapter 2 we considered the system of m equations for n unknowns symbolized by $A\underline{x} = \underline{b}$. In 2.2.4 of the same chapter and in other examples we saw that $A\underline{x}$ is a column with m entries. Its i^{th} entry is (compare equation (3.10)):

   $$A_{i\bullet}\underline{x} = a_{i1}x_1 + a_{i2}x_2 + ... + a_{1n}x_n$$

   Hence,

   $$A\underline{x} = \begin{bmatrix} A_{1\bullet}\underline{x} \\ A_{2\bullet}\underline{x} \\ \vdots \\ A_{m\bullet}\underline{x} \end{bmatrix} \tag{3.22}$$

   We met such a product in 2.2.4 of Chapter 2. The entry in the i^{th} row of $A\underline{x}$ is the 'dot product' $A_{i\bullet}\underline{x}$ of row $A_{i\bullet}$ and column $\underline{x}$.

2. $A\underline{x}$ as a **linear combination of the columns** of A with coefficients $x_1, \ldots, x_n$:

   $$A\underline{x} = A_{\bullet 1}x_1 + A_{\bullet 2}x_2 + ... + A_{\bullet n}x_n \tag{3.23}$$

This view was anticipated in Chapter 2, for example in equations (2.17, (2.19) and (2.25). Like the previous view, it can immediately be seen from the definitions. The following examples should make this clear.

Example 3.1.1 *Let*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

and $\underline{x} \in \Re^{3\times 1}$. *Then*

$$\begin{aligned} A\underline{x} &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} x_2 + \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix} x_3 \end{aligned}$$

So

$$\begin{aligned} &\begin{bmatrix} -3 & 6 & 5 \\ 2 & -4 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ = &\begin{bmatrix} -3x_1 + 6x_2 + 5x_3 \\ 2x_1 - 4x_2 + 7x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} 6 \\ -4 \end{bmatrix} x_2 + \begin{bmatrix} 5 \\ 7 \end{bmatrix} x_3 \end{aligned}$$

Example 3.1.2 *As a more general example, consider the general system of 4 equations for five unknowns, equation (2.11) of the previous chapter. We found there equation (2.12):*

$$A\underline{x} = \begin{bmatrix} A_{1\bullet}\underline{x} \\ A_{2\bullet}\underline{x} \\ A_{3\bullet}\underline{x} \\ A_{4\bullet}\underline{x} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} x_2 + \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix} x_3 + \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix} x_4 + \begin{bmatrix} a_{15} \\ a_{25} \\ a_{35} \\ a_{45} \end{bmatrix} x_5$$

Example 3.1.3 *In the previous example let $\underline{x} = \underline{e}_j$ be the unit vector of equation (3.7) with $n = 5$ entries . Then $A\underline{e}_2 = A_{\bullet 2}$ (the second column of A) since $x_2 = 1$ and $x_1 = x_3 = x_4 = x_5 = 0$. Similarly, for $j = 1, \cdots, 5$, we have $A\underline{e}_j = A_{\bullet j}$.*

3.1.7 $A\underline{e}_j = A_{\bullet j}$ in general

From the above example it is clear that if A is an $m \times n$ matrix and $\underline{e}_j$ is the j^{th} unit vector as in equation (3.7), then

$$A\underline{e}_j = A_{\bullet j} \qquad (j = 1, \cdots, n) \tag{3.24}$$

Equivalently, if I is the $n \times n$ identity matrix,

$$AI_{\bullet j} = A_{\bullet j} \qquad (j = 1, \cdots, n)$$

3.1.8 $A\underline{x}$ as a product of A and $\underline{x}$

Henceforth we will regard $A\underline{x}$ as the **product** of the $m \times n$ matrix A by the $n \times 1$ column matrix $\underline{x}$.

3.1.9 Connection between linear combinations and solving equations, the column space

This was also anticipated in Chapter 2 (for example, in Exercise 32 and in No.2b from Exercise 35.

Solving a system of equations $A\underline{x} = \underline{b}$ for $\underline{x}$ is exactly the same as looking for coefficients x_j so that $\underline{b}$ is a linear combination of the columns of A:

$$A\underline{x} = \underline{b} \text{ if, and only if } \underline{b} = A_{\bullet 1}x_1 + A_{\bullet 2}x_2 + ... + A_{\bullet n}x_n$$

In terms of the concept of 'span' (see item 2 in subsection 3.1.1), this is saying that $\underline{b}$ is in the span of the columns of A:

$$\underline{b} \in sp(A_{\bullet 1}, A_{\bullet 2}, \ldots, A_{\bullet n})$$

Example 3.1.4 *Is $\underline{a} = [1, 5, 2, 7]$ a linear combination of $\underline{u} = [1, 3, 0, 5]$ and $\underline{v} = [0, 1, 1, 1, 1]$? Another way to put it: 'Does $\underline{a}$ depend linearly on $\underline{u}$ and $\underline{v}$?' In other terminology we have been using, 'is it true that $\underline{a} \in sp(\underline{u}, \underline{v})$?'*

Solution:

To answer this, we change the vectors into column vectors (i.e. we use $\underline{u}^T$, $\underline{v}^T$, $\underline{a}^T$) and try to solve the following matrix equation for $\underline{x}$:

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \\ 0 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 2 \\ 7 \end{bmatrix}$$

In other words, we try to express the column on the right as a linear combination of the columns of the 4×2 coefficient matrix. Use the array of detached coefficients:

$$\begin{array}{ccc} 1 & 0 & 1 \\ 3 & 1 & 5 \\ 0 & 1 & 2 \\ 5 & 1 & 7 \end{array}$$

$$\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{array} \quad \begin{array}{l} \\ -3R_1 + R_2 \\ \\ -5R_1 + R_4 \end{array}$$

This is equivalent to

$$\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \quad \begin{array}{l} \\ \\ -R_2 + R_3 \\ -R_2 + R_4 \end{array}$$

Hence $x_2 = 2$ and $x_1 = 1$, and $\underline{a}^T = \underline{u}^T + \underline{v}^T 2$, so $\underline{a} = \underline{u} + 2\underline{v}$ and $\underline{a}$ depends linearly on $\underline{u}$ and $\underline{v}$.

Remark 49 *We are in the habit of solving equations in the form $A\underline{x} = \underline{b}$ (so we are taking linear combinations of columns), but we could just as well have left the vectors as rows: Then we would be taking linear combinations of rows and using elementary* ***column*** *operations.*

Remark 50 *Any result about columns has a corresponding result about rows and vice-versa. This is a fact you will appreciate more and more as we go ahead.*

Example 3.1.5 *Consider from Chapter 2, Example 2.1.4, equation (2.7):*

$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & 4 \\ 2 & 4 & 7 \end{bmatrix} \underline{x} = \begin{bmatrix} -1 \\ -1 \\ 5 \end{bmatrix}$$

We found from the detached array in echelon form (2.20) in Chapter 2 that there is no solution. Thus the right-hand side $\begin{bmatrix} -1 & -1 & 5 \end{bmatrix}^T$ is not a linear combination of the columns of the coefficient matrix A.

Example 3.1.6 *Let*

$$A = \begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix} \tag{3.25}$$

What sort of solutions does the homogeneous system $A\underline{x} = \underline{0}$ have? Is some column of A a linear combination of the other columns of A?

Solution:

Solve the homogeneous linear system

$$\begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\underline{x} = \underline{0}$ is in any case the (trivial) solution. As usual, to find the general solution, use Gauss-reduction to find a row echelon form of the coefficient matrix A:

$$\begin{bmatrix} 2 & 5 & -11 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.26}$$

and the general solution with parameter t is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2t \\ 3t \\ t \end{bmatrix} \tag{3.27}$$

(Check this). If we put $t = 1$, we get

$$\begin{aligned} \begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} &= \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} (-2) + \begin{bmatrix} 5 \\ -6 \\ 7 \end{bmatrix} 3 + \begin{bmatrix} -11 \\ 12 \\ -13 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned} \tag{3.28}$$

Since each coefficient -2, 3 and 1 is non-zero, every column of A is a linearly dependent on the other two columns. For example,

$$A_{\bullet 1} = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ -6 \\ 7 \end{bmatrix} \frac{3}{2} + \begin{bmatrix} -11 \\ 12 \\ -13 \end{bmatrix} \frac{1}{2}$$

Example 3.1.7 *Let*

$$A = \begin{bmatrix} 2 & 5 & 1 \\ -3 & -6 & 0 \\ 4 & 7 & -2 \end{bmatrix} \tag{3.29}$$

Can we find a non-trivial solution $\underline{x}$ to $A\underline{x} = \underline{0}$? Is any column of A linearly dependent on the other two?

Solution:

Reduce A to row-echelon form:

$$\begin{matrix} 2 & 5 & 1 \\ 0 & \frac{3}{2} & \frac{3}{2} \\ 0 & -3 & -4 \end{matrix} \quad \begin{matrix} \\ \frac{3}{2}R_1 + R_2 \\ (-2)\,R_1 + R_3 \end{matrix} \quad , \quad \begin{matrix} 2 & 5 & 1 \\ 0 & \frac{3}{2} & \frac{3}{2} \\ 0 & 0 & -1 \end{matrix} \quad \begin{matrix} \\ \\ 2R_2 + R_3 \end{matrix} \quad , \quad \begin{matrix} 2 & 5 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{matrix} \quad \begin{matrix} \\ \frac{2}{3}R_2 \\ (-1)\,R_3 \end{matrix}$$

So a row EF for A is

$$C = \begin{bmatrix} 2 & 5 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus $A\underline{x} = \underline{0}$ has only the trivial solution

$$\underline{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Hence no column of A can depend linearly on the other two. For example, should we have $A_{\bullet 1} = A_{\bullet 2}\beta + A_{\bullet 3}\gamma$, then

$$\underline{x} = \begin{bmatrix} 1 \\ -\beta \\ -\gamma \end{bmatrix}$$

would be a non-trivial solution to $A\underline{x} = \underline{0}$ because $1 \neq 0$.

Remark 51 *Questions of the sort "which columns of A are linearly dependent on other columns of A?" are further explored below in 3.4.22.*
When $A\underline{x} = \underline{0}$ has only the trivial solution $\underline{x} = \underline{0}$ we will call the columns of A linearly independent. See subsection 3.4.1.

3.1.10 Fundamental properties of the product $A\underline{x}$

Let A be an $m \times n$ matrix and suppose $\underline{u}_1$, $\underline{u}_2$ are vectors in $\Re^n = \Re^{n\times 1}$ and s_1, s_2 are scalars. Then

$$A\left(\underline{u}_1 s_1 + \underline{u}_2 s_2\right) = \left(A\underline{u}_1\right) s_1 + \left(A\underline{u}_2\right) s_2 \tag{3.30}$$

Proof

Consider the i^{th} entry of the left-hand side of (3.30). By equation (3.16) this is

$$A_{i\bullet}\left(\underline{u}_1 s_1 + \underline{u}_2 s_2\right) = \left(A_{i\bullet}\underline{u}_1\right) s_1 + \left(A_{i\bullet}\underline{u}_2\right) s_2$$

As $\left(A_{i\bullet}\underline{u}_1\right) s_1 + \left(A_{i\bullet}\underline{u}_2\right) s_2$ is by definition the i^{th} entry of the right-hand side of equation (3.30), the result follows.
More generally, equation (3.30) easily extends to the following general result:

Let A be an $m \times n$ matrix and suppose $\underline{u}_1$, $\underline{u}_2$, ..,$\underline{u}_k$ are vectors in $\Re^n = \Re^{n\times 1}$ and s_1, s_2, ..,.s_k are scalars. Then

$$A\left(\underline{u}_1 s_1 + \underline{u}_2 s_2 + ... + \underline{u}_k s_k\right) = \left(A\underline{u}_1\right) s_1 + \left(A\underline{u}_2\right) s_2 + ... + \left(A\underline{u}_k\right) s_k \tag{3.31}$$

Exercise 52 *1. Find $I\underline{x}$ if $I = I_n$ is the $n \times n$ identity matrix as in equation (3.8) and $\underline{x} \in \Re^{n\times 1}$.*

2. Let A be $m \times n$. Show that $\beta A_{2\bullet} + A_{1\bullet}$ is a linear combination of $A_{2\bullet}, \cdots, A_{m\bullet}$ if, and only if, $A_{1\bullet}$ is a linear combination of $A_{2\bullet}, \cdots, A_{m\bullet}$.

3. Suppose that $\underline{v}_1$, $\underline{v}_2$, ... , $\underline{v}_k$ are solutions to a system $A\underline{x} = \underline{b}$ of m equations in n unknowns $x_1, \ldots, x_n$. Let $\underline{w} = \underline{v}_1\alpha_1 + \underline{v}_2\alpha_2 + \cdots + \underline{v}_k\alpha_k$ be a linear combination of these vectors.

(a) Show that if the system is homogeneous i.e. $\underline{b} = \underline{0}$, then $\underline{w}$ is also a solution, i.e. $A\underline{w} = \underline{0}$

***Solution**:*
$A\underline{w} = A\left(\underline{v}_1\alpha_1 + \underline{v}_2\alpha_2 + \cdots + \underline{v}_k\alpha_k\right) = \left(A\underline{v}_1\right)\alpha_1 + \left(A\underline{v}_2\right)\alpha_2 + \cdots + \left(A\underline{v}_k\right)\alpha_k = \underline{0}\alpha_1 + \cdots \underline{0}\alpha_k = \underline{0}$

(b) Find a simple example to show that (a) may fail if $\underline{b} \neq \underline{0}$.

(c) If $\alpha_1 + \alpha_2 + \cdots + \alpha_k = 1$ show that $A\underline{w} = \underline{b}$.

4. (a) Consider Exercise 35 of Chapter 2, Numbers 3, 4a, 5a and 6, 11a and 11b. These require solutions to a matrix equation $A\underline{x} = \underline{b}$. For which of these is $\underline{b}$ a linear combination of the columns of A?
Consider those for which $\underline{b}$ is a linear combination of the columns of A. For which of these are the coefficients in the linear combination unique? For those which are not, express $\underline{b}$ in two ways as linear combinations of the columns of A.

*(b) * In Problem 13 from Exercise 35 of Chapter 2. For which values of a, b, c, d is $\begin{bmatrix} a & b & c & d \end{bmatrix}^T$ a linear combination of the columns of A?*

5. Prove the following theorem, which was anticipated in No.10 of Exercise 35 in Chapter 2.

Let $A\underline{x} = \underline{b}$ represent m equations in n unknowns. Suppose that the system has at least

one particular solution, say $\underline{p}$. Then if $A\underline{v} = \underline{0}$ so is $\underline{x} = \underline{p} + \underline{v}$ a solution and all solutions to $A\underline{x} = \underline{b}$ have this form.
Conclude that if $m < n$ and $A\underline{x} = \underline{b}$ has at least one solution, then the system has infinitely many solutions. In fact, if there are k non-pivot columns then the general solution contains k independent parameters. (Use 2.2.15 from Chapter 2.)

3.2 Matrices as Mappings

Let A be an $m \times n$ matrix. In what follows $\Re^n$ stands for the set of column vectors with n entries, ie. $\Re^n = \Re^{n\times 1}$. The matrix A converts or **transforms** the vector $\underline{x} \in \Re^n$ into the vector $A\underline{x} \in \Re^m$. In other words, letting $\underline{y} = A\underline{x}$, have a function of the *independent* variable $\underline{x}$ with $\underline{y} = A\underline{x}$ being the *dependent* variable. In the language of set-theory, A is a **function** or **transformation** or **mapping** with domain $\Re^n$ and range in the set $\Re^m$:

$$A : \Re^n \to \Re^m \tag{3.32}$$

A **maps** the vector $\underline{x}$ to the vector $\underline{y} = A\underline{x}$. Symbolically,

$$\underline{x} \xrightarrow{\quad A \quad} A\underline{x} \quad \text{or} \quad \underline{x} \to A\underline{x}$$

3.2.1 The range

The **range** of $\underline{x} \to A\underline{x}$ is the set of vectors $\underline{y}$ such that $A\underline{x} = \underline{y}$ for some $\underline{x}$. In other words, the range is just the set of all linear combinations of the columns of A. The range is therefore the column space of A (see item 2 in subsections 3.1.1 and 3.1.9).
See also section 3.6, where some important properties of subspaces are developed.

Remark 53 *If $\underline{c}$ is a fixed column vector with m entries then $\underline{x} \to \underline{c} + A\underline{x}$ is also a mapping from $\Re^n$ to $\Re^m$. However, we will concentrate mainly on the case $\underline{c} = \underline{0}$ (linear transformations).*

3.2.2 Solving equations is the inverse process of the mapping $\underline{x} \to A\underline{x}$

Given the mapping (3.32) and a vector $\underline{y} \in \Re^m$, we can ask: Is there an $\underline{x} \in \Re^n$ such that $\underline{y} = A\underline{x}$? This is precisely the problem of searching for a solution $\underline{x}$ to the system of equations $\underline{y} = A\underline{x}$, as discussed in 3.1.9. In Chapter 2 we found that some systems have a solution, others not.

Example 3.2.1 *Consider the Cartesian plane $\Re^2$. In equation (1.27) of Chapter 1 we found the reflection of the point (p_1, p_2) in the line ℓ lying in the $x-y$ plane and which passes through the origin making an angle of 30° with the positive $x-$axis:*

$$\underline{p}' = \left(\frac{1}{2}p_1 + \frac{1}{2}\sqrt{3}p_2, \frac{1}{2}\sqrt{3}p_1 - \frac{1}{2}p_2\right)$$

Hence considered as a mapping, the matrix

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\sqrt{3} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$$

transforms the point $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ into its reflection in the above line:

$$\underline{y} = A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}x_1 + \frac{1}{2}\sqrt{3}x_2 \\ \frac{1}{2}\sqrt{3}x_1 - \frac{1}{2}x_2 \end{bmatrix}$$

(We have written $\underline{p}$ and $\underline{p}'$ as columns $\underline{x}$ and $\underline{y}$ respectively). See Figure 3.1. Geometrically, it is clear that the range of this reflection is $\Re^2$. This can be seen algebraically by showing that, whatever the values of y_1 and y_2, the matrix equation

$$\begin{bmatrix} \frac{1}{2}x_1 + \frac{1}{2}\sqrt{3}x_2 \\ \frac{1}{2}\sqrt{3}x_1 - \frac{1}{2}x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

always has a solution for x_1 and x_2. (Show this).

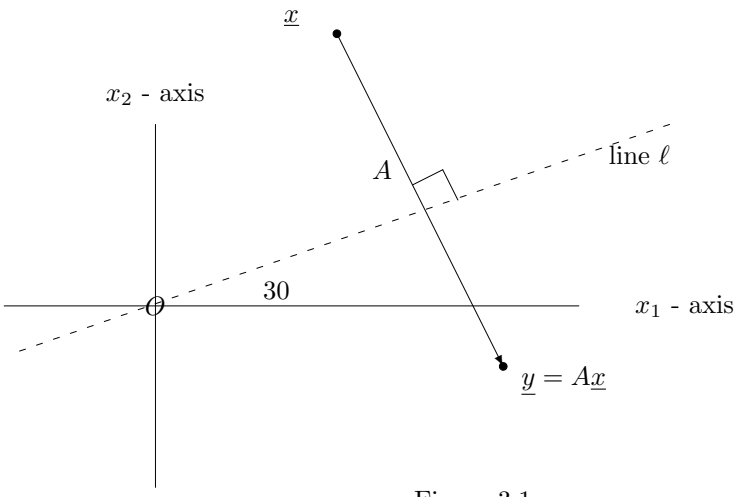


Figure 3.1

Example 3.2.2 *The simpler matrix*

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

represents a reflection in the x_2-axis. The range is again $\Re^2$.

Example 3.2.3 *The matrix*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

represents the mapping that projects the point $\underline{x} = [x_1, x_2, x_3]^T \in \Re^3$ in space onto the point $\underline{y} = A\underline{x} = [x_1, x_2]^T \in \Re^2$ in the $x_1 - x_2$ plane. The range is obviously $\Re^2$.

Example 3.2.4 *Consider a fixed column vector $\underline{a}$ with three entries and non-zero column matrix*

$$A = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

The mapping, $A : \Re^3 \to \Re^1 = \Re$, and $t \to \underline{a} + At$ represents a parametric equation of a straight line in space passing through the point $\underline{a}$, as in Chapter 1, 1.2. The only difference is that here we are using columns instead of rows.

Example 3.2.5 *Consider the matrix*

$$A = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \\ u_3 & v_3 \end{bmatrix}$$

in which the columns are non-zero and non-parallel. The mapping

$$\begin{bmatrix} s \\ t \end{bmatrix} \to A \begin{bmatrix} s \\ t \end{bmatrix}$$

now represents a generic equation of a plane going through the origin.
See Chapter 1, subsection 1.3.1.

Example 3.2.6 *The matrix*

$$A = \begin{bmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -4 \end{bmatrix}$$

*transforms the point $\underline{x} \in \Re^3$ into the point $\underline{y} = A\underline{x} = -4\underline{x} \in \Re^3$. This represents a **stretching** $\underline{x} \to 4\underline{x}$ of $\underline{x}$ by a factor of 4 followed by an **inversion** $\underline{x} \to -\underline{x}$ in the origin O.*

Example 3.2.7 *The matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ as mapping of $\Re^2$ to $\Re^2$ is*

$$\underline{y} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

*and represents a **rotation** of the point with position vector $\underline{x}$ through $90°$ (anticlockwise looking down on the plane). Convince yourselves of this by looking at a few values of $\underline{x}$, but we will return to this below in 3.3.11.*

Example 3.2.8 *The matrix $A = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}$, interpreted as a mapping, leaves x_2 fixed and moves x_1 to $x_1 + \alpha x_2$. It is known as a shear transformation.*

$$A\underline{x} = \begin{bmatrix} x_1 + \alpha x_2 \\ x_2 \end{bmatrix}$$

See Figure 3.2 where $\alpha > 0$. The box $OBCD$ is transformed into the quadrilateral $OEFD$.
The range is again $\Re^2$, as can very easily be seen. (Exercise 56 No.3).

3.2.3 $A\underline{x} = B\underline{x}$ for all $\underline{x}$ if, and only if, $A = B$

It is obvious that if $A = B$ then $A\underline{x} = B\underline{x}$ for all $\underline{x}$.
Conversely, if $A\underline{x} = B\underline{x}$ for $\underline{x} \in \Re^n$ then A and B must have the same number n of columns and also the same number m of rows (the number of rows in $A\underline{x} = B\underline{x}$). Furthermore, if $\underline{e}_j \in \Re^n$ is the j^{th} unit vector, then by equation (3.24), $A_{\bullet j} = A\underline{e}_j = B\underline{e}_j = B_{\bullet j}$ for $j = 1, \cdots, n$. Hence A and B have the same columns and so $A = B$.

3.3 Linear Transformations

Because of equation (3.31), the mapping (3.32) is said to be **linear** and is called a *linear transformation.*

In abstract terms, a mapping

$$T : \Re^n \to \Re^m \tag{3.33}$$

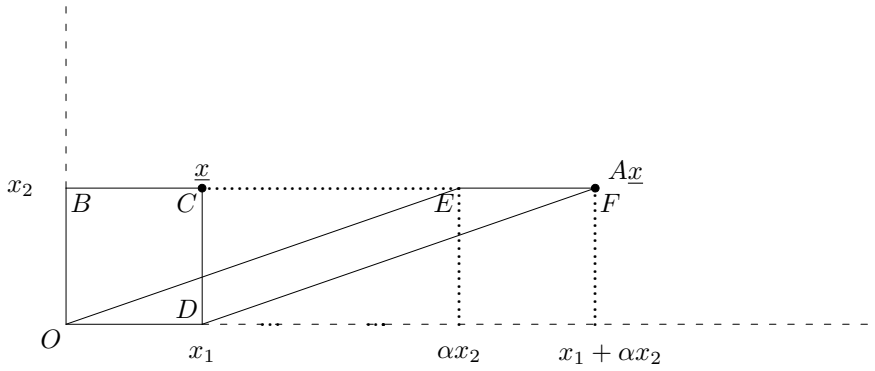


Figure 3.2

is called a **linear transformation** if for all $\underline{u}_1$ and $\underline{u}_2$ in $\Re^n$ and all scalars s_1 and s_2 it follows that

$$T\left(\underline{u}_1 s_1 + \underline{u}_2 s_2\right) = T\left(\underline{u}_1\right) s_1 + T\left(\underline{u}_2\right) s_2 \tag{3.34}$$

An equivalent definition is the following.

The mapping (3.33) is a linear transformation if, and only if, for all $\underline{u}$ and $\underline{v}$ in $\Re^n$ and all scalars s,

$$T\left(\underline{u} + \underline{v}\right) = T\left(\underline{u}\right) + T\left(\underline{v}\right) \tag{3.35}$$

$$T\left(\underline{u}s\right) = T\left(\underline{u}\right) s \tag{3.36}$$

For it is obvious that if (3.34) holds then so will (3.35) and (3.36). On the other hand, suppose that these last two conditions hold for the mapping T. Then for any $\underline{u}_1$ and $\underline{u}_2$ scalars s_1 and s_2,

$$T\left(\underline{u}_1 s_1 + \underline{u}_2 s_2\right) = T\left(\underline{u}_1 s_1\right) + T\left(\underline{u}_2 s_2\right) = T\left(\underline{u}_1\right) s_1 + T\left(\underline{u}_2\right) s_2$$

Remark 54 *By putting $s = 0$ in (3.36) we see that a necessary condition for T to be linear is that it maps the zero vector of $\Re^n$ to the zero vector of $\Re^m$. Most transformations are decidedly non-linear. Consider, for example, $T : \Re^1 \to \Re^1$ given by $T(x) = x^2$. Here $T(0) = 0$ but T fails (3.35), since (e.g.) $T(1+1) = 2^2 \neq T(1) + T(1) = 2$.*

3.3.1 An extension of equation (3.34)

The result expressed by equation (3.34) easily extends to the following:

Let T be a linear transformation and suppose vectors $\underline{u}_1$, $\underline{u}_2$, ..,$\underline{u}_k$ in $\Re^n$ are given as well as scalars s_1, s_2, ...s_k, then

$$T\left(\underline{u}_1 s_1 + \underline{u}_2 s_2 + ... + \underline{u}_k s_k\right) = T\left(\underline{u}_1\right) s_1 + T\left(\underline{u}_2\right) s_2 + ... + T\left(\underline{u}_k\right) s_k \tag{3.37}$$

We use this result to show that

3.3.2 A linear transformation $T : \Re^n \to \Re^m$ is defined by a unique matrix mapping $\underline{x} \to A\underline{x}$

Proof:

We have seen that the mapping $\underline{x} \to A\underline{x}$ is a linear transformation.
Conversely, let $T : \Re^n - \Re^m$ be a linear transformation. We show that there is a unique matrix A such that $T(\underline{x}) = A\underline{x}$ for all $\underline{x} \in \Re^n$.

Let $\underline{e}_j$ be one of the n unit vectors (3.7). By assumption, $T\left(\underline{e}_j\right)$ is a column vector with m entries. Since $\underline{x} \in \Re^n$ has the expression $\underline{x} = \underline{e}_1 x_1 + \underline{e}_2 x_2 + ... + \underline{e}_n x_n$, we deduce from equation (3.37) that

$$T(\underline{x}) = T(\underline{e}_1)\, x_1 + T(\underline{e}_2)\, x_2 + ... + T(\underline{e}_n)\, x_n \tag{3.38}$$

In other words, by equation (3.23),

$$T(\underline{x}) = A\underline{x} \qquad (\underline{x} \in \Re^n) \tag{3.39}$$

Here the matrix A is defined by $A_{\bullet j} = T\left(\underline{e}_j\right)$ for $j = 1, \cdots, n$. This is saying that A is the $m \times n$ matrix

$$A = \left[\begin{array}{cccc} T(\underline{e}_1) & T(\underline{e}_2) & ... & T(\underline{e}_n) \end{array} \right]$$

The matrix A in (3.39) is obviously unique.

3.3.3 A linear transformation T is completely determined by its effect $T\left(\underline{e}_j\right)$ on the unit vectors $\underline{e}_j$

This is just a restatement of equation (3.38).

Remark 55 *Since a linear transformation is just a mapping determined by a matrix, why bother with the concept 'linear transformation'? One reason is that our intuition often suggests that a transformation is linear.*

Example 3.3.1 *In equation 1.14 of Chapter 1 we found the projection Q of a point $P = (p_1, p_2, p_3)$ on the line $t(-2, -1, 1)$:*

$$\underline{q} = \left(\frac{2}{3}p_1 + \frac{1}{3}p_2 - \frac{1}{3}p_3, \frac{1}{3}p_1 + \frac{1}{6}p_2 - \frac{1}{6}p_3, -\frac{1}{3}p_1 - \frac{1}{6}p_2 + \frac{1}{6}p_3 \right)$$

Our geometric intuition strongly suggests that this should represent a linear transformation. Prove this by finding the matrix determining it.

Solution:

Writing $x_1 = p_1$, $x_2 = p_2$, $x_3 = p_3$ and using column vectors, the above projection formula becomes

$$\left[\begin{array}{c} q_1 \\ q_2 \\ q_3 \end{array} \right] = \left[\begin{array}{ccc} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{3} & -\frac{1}{6} & \frac{1}{6} \end{array} \right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right]$$

This is the mapping $\underline{x} \to \underline{q} = A\underline{x}$ and is indeed a linear transformation.

Example 3.3.2 *Find a matrix representation of the mapping which sends the point P to the foot of the perpendicular from P to the plane $x_1 + 2x_2 + x_3 = -4$. Show that this mapping is not linear but that the projection of P on the plane $x_1 + 2x_2 + x_3 = 0$ is a linear transformation. Find the matrix that reflects in this plane.*

Solution:

From equation (1.22) in Chapter 1, with $P = (p_1, p_2, p_3)$ we found for the foot Q

$$\underline{q} = (p_1, p_2, p_3) - \frac{4 + p_1 + 2p_2 + p_3}{6}(1, 2, 1)$$

Letting $x_1 = p_1$, $x_2 = p_2$, $x_3 = p_3$, and again using column vectors, this becomes

$$\begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} \frac{2}{3} \\ \frac{4}{3} \\ \frac{2}{3} \end{bmatrix} + \begin{bmatrix} -\frac{x_1+2x_2+x_3}{6} \\ -2\frac{x_1+2x_2+x_3}{6} \\ -\frac{x_1+2x_2+x_3}{6} \end{bmatrix}$$

This cannot be linear since $\underline{x} = \underline{0}$ does not map to $\underline{q} = \underline{0}$.

However, from equation (1.23) we found the formula for the projection Q of $P = (p_1, p_2, p_3)$ on the plane $x_1 + 2x_2 + x_3 = 0$ to be

$$\underline{q} = (p_1, p_2, p_3) - \frac{p_1 + 2p_2 + p_3}{6}(1, 2, 1)$$

In matrix form with $P = \underline{x}$, this reads

$$\begin{aligned} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} &= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} -\frac{x_1+2x_2+x_3}{6} \\ -2\frac{x_1+2x_2+x_3}{6} \\ -\frac{x_1+2x_2+x_3}{6} \end{bmatrix} \\ &= \begin{bmatrix} \frac{5}{6}x_1 - \frac{1}{3}x_2 - \frac{1}{6}x_3 \\ -\frac{1}{3}x_1 + \frac{1}{3}x_2 - \frac{1}{3}x_3 \\ -\frac{1}{6}x_1 - \frac{1}{3}x_2 + \frac{5}{6}x_3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned} \tag{3.40}$$

The matrix that reflects in the plane $x_1 + 2x_2 + x_3 = 0$ is

$$2\begin{bmatrix} \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

This also defines a linear transformation.

Exercise 56 ***Partial Solutions***

1. *Find a matrix formula for the foot Q of the perpendicular from $P = \underline{x} \in \Re^3$ to the plane $2x + z - 3 = 0$. Do the same for the plane $2x + z = 0$, showing in the latter case that we get a linear transformation. Write down the matrix A of this transformation, as well as the matrix that reflects in the plane $2x + z = 0$.*

   Hint 57 *Refer to 1d from Exercise 16 in Chapter 1. There you found the foot Q of the perpendicular from (p_1, p_2, p_3) to the plane $2x + z - 3 = 0$:*

   $$\underline{q} = \left(\frac{1}{5}p_1 - \frac{2}{5}p_3 + \frac{6}{5}, p_2, \frac{4}{5}p_3 - \frac{2}{5}p_1 + \frac{3}{5}\right)$$

   Now use appropriate columns in place of row vectors.

2. *(a) Consider equation (1.13) in subsection 1.2.5 of Chapter 1 for the projection of a point on a line ℓ that passes through the origin. Using column vectors find a matrix A such that the foot Q of the perpendicular from the point $\underline{x}$ is given by $\underline{q} = A\underline{x}$.*

(b) *Consider equation (1.24) from Exercise 16 in Chapter 1. Assume that the plane goes through the origin. Using column vectors find a matrix equation* $\underline{q} = A\underline{x}$ *for the foot* Q *of the perpendicular from the point* $\underline{x}$ *onto the plane (the projection of* P *onto the plane).*

The matrices in 2a and 2b are called ***projection matrices****.*

(c) *Specialize (a) in case the transformation is restricted to the Cartesian plane, i.e. is from* $\Re^2$ *to* $\Re^2$*.*

3. *Find the range of the shear transformation (Example 3.2.8) and prove that your statement is correct.*

4. *Use Exercise 20 from Chapter 1 to find linear transformations describing the following:*

(a) *The reflection in a line that passes through the origin.*

Remark 58 *Observe that this matrix actually represents a rotation through* π *radians about the given line. Contrast this with the situation when we restrict ourselves to the Cartesian plane in Exercise 63, No.8, where a reflection cannot be a rotation.*

(b) *The reflection in a plane that passes through the origin.*

(c) *Specialize (a) in case the transformation is restricted to the Cartesian plane, i.e. is from* $\Re^2$ *to* $\Re^2$*.*

(d) *What do you think the ranges of these mappings are?*

5. *In the* x_1 *-* x_2 *plane let* A *be the rotation matrix of Example 3.2.7 and suppose* B *is the* 2×2 *matrix that reflects in the* x_1-*axis.*

(a) *Find matrices* C *and* D *such that for all vectors* $\underline{x} \in \Re^2$ *we have* $C\underline{x} = B(A\underline{x})$ *and* $D\underline{x} = A(B\underline{x})$*.*

(b) *Show that* C *and* D *are reflection matrices and describe them geometrically.*

Remark 59 *In the next section, we will regard* C *as the product of* B *and* A*,* ***in that order****:* $C = BA$*. Similarly,* $D = AB$*. See Example 3.3.3 and notice that* $AB \neq BA$*.*

6. *Let* A *be one of the projection matrices from the previous exercises. Show geometrically why you expect that* $A(A\underline{x}) = A\underline{x}$ *for all points* $\underline{x}$*. Now suppose that* A *is one of the reflection matrices. What should* $A(A\underline{x})$ *be? See also Exercise 63, No.10 below.*

3.3.4 The product AB of two matrices

Let A be an $m \times k$ matrix and B a $k \times n$ matrix. As linear transformations,

$$B : \Re^n \to \Re^k \quad \text{and} \quad A : \Re^k \to \Re^m.$$

Consequently, for $\underline{x} \in \Re^n$

$$\underline{x} \to A(B\underline{x}) \tag{3.41}$$

defines a transformation from $\Re^n$ to $\Re^m$. See Figure 3.3. Symbolically

$$\underline{x} \xrightarrow{\quad B \quad} B\underline{x} \xrightarrow{\quad A \quad} A(B\underline{x})$$

It is essential to observe that (3.41) only makes sense if **the number of columns of** A equals **the number of rows of** B (here both are k).

Remark 60 *We have to read this as:* ***first*** *apply* B *to* $\underline{x}$*,* ***then*** *apply* A *to the result* $B\underline{x}$*. This "backward" reading is due to our traditional functional notation: For ordinary functions* f *and* g *we write the composition of* f *and* g *as* $(f \circ g)(x) = f(g(x))$*.*

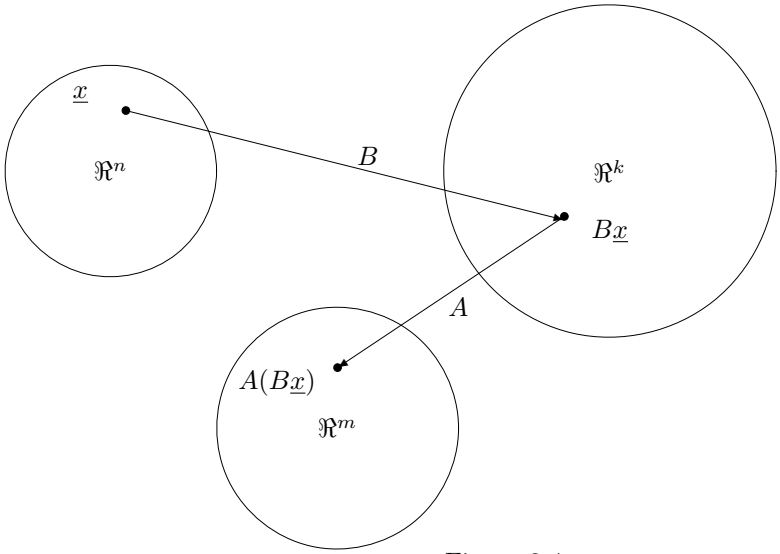


Figure 3.4

We will define the product $P = AB$ in such a way that

$$P\underline{x} = A\left(B\underline{x}\right) \qquad (\underline{x} \in \Re^n) \tag{3.42}$$

In other words, so that AB defines the **composition** of the transformations A and B.

3.3.5 There is one, and only one, matrix P satisfying equation (3.42)

Proof

Suppose (3.42) holds. Put $\underline{x} = \underline{e}_j$, one of the n unit vectors (3.7). Then

$$P_{\bullet j} = P\underline{e}_j = A\left(B\underline{e}_j\right) = AB_{\bullet j} \qquad (j = 1, \cdots, n) \tag{3.43}$$

In other words, for (3.42) to hold (3.43) is the only choice for P. On the other hand, (3.41) is a linear transformation: See No.15 in Exercise 63). Therefore its matrix can only have columns (3.43). We therefore define the matrix AB as follows:

3.3.6 Definition

If A is $m \times k$ and B is $k \times n$, then

$$AB = \left[\begin{array}{cccc} AB_{\bullet 1} & AB_{\bullet 2} & \cdots & AB_{\bullet n} \end{array}\right] \tag{3.44}$$

Equivalently,

$$[AB]_{\bullet j} = AB_{\bullet j} \quad (j = 1, \cdots, n) \tag{3.45}$$

With this notation So writing AB for P we rewrite (3.42) as

$$(AB)\,\underline{x} = A\left(B\underline{x}\right) \qquad (\underline{x} \in \Re^n)$$

3.3.7 Alternative more symmetric definition of AB

Equation (3.44) seems biased towards columns, but that this is not the case can be seen by looking at the entries of $(AB)_{\bullet j}$. Using (3.22) with $\underline{x} = B_{\bullet j}$ and the notation of equation (3.1),

$$[AB]_{ij} = A_{i\bullet}B_{\bullet j} \quad (i = 1, \cdots, m;\ \ j = 1, \cdots, n) \tag{3.46}$$

The entry in row i and column j of AB is the product of row $A_{i\bullet}$ and column $B_{\bullet j}$. Equation (3.46) can just as well be taken as the definition of AB and in fact is the standard definition. This equation was anticipated in Exercise 47, No.10a above.

Whatever is true for columns (rows) has its exact counterpart for rows (columns). From (3.46) we see that

$$[AB]_{i\bullet} = A_{i\bullet}B \quad (i = 1, \cdots, m) \tag{3.47}$$

or more fully,

$$AB = \begin{bmatrix} A_{1\bullet}B \\ A_{2\bullet}B \\ \vdots \\ A_{m\bullet}B \end{bmatrix}$$

Expanding (3.46), we get using the usual convention for matrix entries and $\sum$-notation,

$$[AB]_{ij} = \sum_{r=1}^{k} a_{ir}b_{rj} = a_{i1}b_{1j} + \cdots + a_{ik}b_{kj}$$

As already remarked, this equation only makes sense if the number of columns of A equals the number of rows of B. We may write

$$(AB)_{m\times n} = A_{m\times k}B_{k\times n}$$

Example 3.3.3 *Let B be the reflection matrix of Example 3.2.1 and A that of the rotation matrix A of Example 3.2.7. Find the matrix of (i) first rotating then reflecting and (ii) the other way round, first reflecting then rotating.*
See No.5 in Exercise 56.

Solution (i)
If we first rotate through 90° then reflect in the x_1-axis we get the product

$$BA = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 0 & -1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} & \begin{bmatrix} 0 & -1 \end{bmatrix}\begin{bmatrix} -1 \\ 0 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Solution (ii)
If we first reflect and then rotate we obtain

$$AB = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 & -1 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Example 3.3.4 *In Exercise 47, No.10a,*

$$A = \begin{bmatrix} 2 & -9 \\ 3 & 1 \end{bmatrix},\ \ B = \begin{bmatrix} 8 & 2 & -1 \\ -5 & 4 & 7 \end{bmatrix}$$

Find AB.

Solution:

From (3.46),

$$\begin{aligned} AB &= \begin{bmatrix} \begin{bmatrix} 2 & -9 \end{bmatrix}\begin{bmatrix} 8 \\ -5 \end{bmatrix} & \begin{bmatrix} 2 & -9 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} & \begin{bmatrix} 2 & -9 \end{bmatrix}\begin{bmatrix} -1 \\ 7 \end{bmatrix} \\ \begin{bmatrix} 3 & 1 \end{bmatrix}\begin{bmatrix} 8 \\ -5 \end{bmatrix} & \begin{bmatrix} 3 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} & \begin{bmatrix} 3 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 7 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 61 & -32 & -65 \\ 19 & 10 & 4 \end{bmatrix} \end{aligned}$$

Example 3.3.5 *Find $\underline{y}B$, where $\underline{y} = \begin{bmatrix} y_1 & y_2 \end{bmatrix}$ and B is as in the previous example:.*

$$\begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} 8 & 2 & -1 \\ -5 & 4 & 7 \end{bmatrix} \tag{3.48}$$

Solution:

$$\begin{aligned} \underline{y}B &= \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} 8 & 2 & -1 \\ -5 & 4 & 7 \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} 8 \\ -5 \end{bmatrix} & \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} & \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} -1 \\ 7 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 8y_1 - 5y_2 & 2y_1 + 4y_2 & -y_1 + 7y_2 \end{bmatrix} \end{aligned}$$

Example 3.3.6 *Express equation (3.48) as a linear combination of rows of B.*

Solution:

In (3.23) we found that the product $A\underline{x}$ is a linear combination of the **columns** of A with coefficients x_k. In exactly the same way we find the product (3.48) to be a linear combination of the **rows** of B with coefficients y_1 and y_2:

$$\begin{aligned} & y_1 B_{1\bullet} + y_2 B_{2\bullet} \\ = \; & y_1 \begin{bmatrix} 8 & 2 & -1 \end{bmatrix} + y_2 \begin{bmatrix} -5 & 4 & 7 \end{bmatrix} \\ = \; & \begin{bmatrix} 8y_1 - 5y_2 & 2y_1 + 4y_2 & -y_1 + 7y_2 \end{bmatrix} \\ = \; & \underline{y}B \end{aligned}$$

3.3.8 For a row $\underline{y}$, the product $\underline{y}B$ is a linear combination of the rows of B with coefficients y_i

Let B be a $k \times n$ matrix and $\underline{y} = [y_1, \cdots, y_k]$ a row vector with k entries. Then

$$\underline{y}B = y_1 B_{1\bullet} + \cdots + y_k B_{k\bullet} \tag{3.49}$$

The reasoning is just as in the case of (3.23) except that now we are multiplying on the *left* of B by a *row* $\underline{y}$ and taking a linear combination of *rows* of B. Like equation (3.22) we have

$$\underline{y}B = \begin{bmatrix} \underline{y}B_{\bullet 1} & \underline{y}B_{\bullet 2} & \cdots & \underline{y}B_{\bullet n} \end{bmatrix}$$

which is just a special case of equation (3.44).

3.3.9 The rows of AB as linear combinations of rows of B and the columns of AB as linear combinations of columns of A

Let $A = [a_{ir}]$ be $m \times k$ and $B = [b_{rj}]$ be $k \times n$. In the product (3.46), we have

1.
$$[AB]_{i\bullet} = a_{i1}B_{1\bullet} + a_{i2}B_{2\bullet} + \cdots + a_{ik}B_{k\bullet} \quad (i = 1, \cdots, m) \tag{3.50}$$

2.
$$[AB]_{\bullet j} = A_{\bullet 1}b_{1j} + A_{\bullet 2}b_{2j} + \cdots + A_{\bullet k}b_{kj} \tag{3.51}$$

To see equation (3.50), use equations (3.47) and (3.49) with $\underline{y} = A_{i\bullet}$.
In words: "Row i of AB is a linear combination of the rows of B with coefficients taken from row i of A".
To see equation (3.51) use equations (3.45) and our earlier result equation (3.23) with $\underline{x} = B_{\bullet j}$.
In words: "Column j of AB is a linear combination of the columns of A with coefficients taken from column j of B".
The above examples illustrate these statements, as do the following.

Example 3.3.7 *Find AB in two ways: (i) in terms of linear combinations of rows of B and (ii) in terms of linear combinations of columns of A, given that*

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 4 \\ -7 & 0 \end{bmatrix}$$

Solution:

(i) In terms of the rows of AB:

$$\begin{aligned} AB &= \begin{bmatrix} A_{1\bullet}B \\ A_{2\bullet}B \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}\begin{bmatrix} 5 & 4 \\ -7 & 0 \end{bmatrix} \\ \begin{bmatrix} -1 & 3 \end{bmatrix}\begin{bmatrix} 5 & 4 \\ -7 & 0 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 2\begin{bmatrix} 5 & 4 \end{bmatrix} + 1\begin{bmatrix} -7 & 0 \end{bmatrix} \\ (-1)\begin{bmatrix} 5 & 4 \end{bmatrix} + 3\begin{bmatrix} -7 & 0 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 3 & 8 \\ -26 & -4 \end{bmatrix} \end{aligned}$$

(ii) In terms of the columns of AB:

$$\begin{aligned} AB &= \begin{bmatrix} AB_{\bullet 1} & AB_{\bullet 2} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 5 \\ -7 \end{bmatrix} & \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ 0 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix}5 + \begin{bmatrix} 1 \\ 3 \end{bmatrix}(-7) & \begin{bmatrix} 2 \\ -1 \end{bmatrix}4 + \begin{bmatrix} 1 \\ 3 \end{bmatrix}0 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 8 \\ -26 & -4 \end{bmatrix} \end{aligned}$$

Example 3.3.8 *If A is a 3×3 matrix, find matrices B and C such that*

$$BA = \begin{bmatrix} -5A_{2\bullet} + 67A_{1\bullet} + 92A_{3\bullet} \\ 9A_{3\bullet} - 4A_{1\bullet} \\ 8A_{2\bullet} + 21A_{1\bullet} \end{bmatrix}$$

and

$$AC = \begin{bmatrix} 7A_{\bullet 1} - 37A_{\bullet 2} & 19A_{\bullet 2} + 17A_{\bullet 1} - 7A_{\bullet 3} \end{bmatrix}$$

(AC has two columns).

Solution:

$$B = \begin{bmatrix} 67 & -5 & 92 \\ -4 & 0 & 9 \\ 21 & 8 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 7 & 17 \\ -37 & 19 \\ 0 & -7 \end{bmatrix}$$

Example 3.3.9 *What is the matrix of the transformation $T : \Re^3 \to \Re^3$ that first projects onto the plane $x_1 + 2x_2 + x_3 = 0$ as in (3.40) and then projects onto the $x_1 - x_2$ plane as in Example 3.2.3?*

Solution:

The required matrix is the product

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

3.3.10 The Associative law: $(AB)\,C = A\,(BC)$

If A is $m \times k$ and B is $k \times n$, our whole inspiration (3.42) for the definition of the product $P = AB$ is that $(AB)\,\underline{x} = A\,(B\underline{x})$ should hold for all $\underline{x} \in \Re^n$. This is a special case of the associative law for multiplication of matrices, but in fact gives us the general law directly.

Statement and proof of the associative law

Let A be an $m \times k$ matrix, B a $k \times n$ matrix and C an $n \times p$ matrix. Then

$$(AB)\,C = A\,(BC) \tag{3.52}$$

Considered as mappings, both sides of (3.52) have the same meaning: First apply C, then B, then A. Therefore both sides are equal as matrix products.

Powers of A if A is square. Let A be $m \times m$. Then $(AA)A$ and $A(AA)$ are both defined and are equal by (3.52), so we can write A^3 for this product. Similarly, if $n \geqslant 1$ is an integer we have $A^n = AA \cdots A$ (n times), where the bracketing does not matter. By convention, $A^0 = I_m$ (the $m \times m$ identity matrix).

Remark 61 *We have given a motivated and especially simple proof of the associative law (3.52). The usual proof given in textbooks is the following:*

$$\begin{aligned} [A\,(BC)]_{ij} &= \sum_{r=1}^{k} [A]_{ir}\,[BC]_{rj} = \sum_{r=1}^{k} [A]_{ir} \left(\sum_{s=1}^{n} [B]_{rs}\,[C]_{sj} \right) \\ &= \sum_{r=1}^{k} \sum_{s=1}^{n} [A]_{ir}\,[B]_{rs}\,[C]_{sj} = \sum_{s=1}^{n} \left(\sum_{r=1}^{k} [A]_{ir}\,[B]_{rs} \right) [C]_{sj} \\ &= \sum_{s=1}^{n} [AB]_{is}\,[C]_{sj} = [(AB)\,C]_{ij} \end{aligned}$$

3.3.11 Rotations are linear transformations

.

Let X be a point in the Cartesian plane $\Re^2 = \Re^{2\times 1}$ and suppose that R rotates $\underline{OX} = \underline{x}$ through the angle θ about the origin *anticlockwise* looking down the x_3- axis towards O. This is a *clockwise* rotation of θ looking from the origin O along the positive x_3-axis. Let $R(\underline{x})$ be the effect of R on $\underline{x}$.

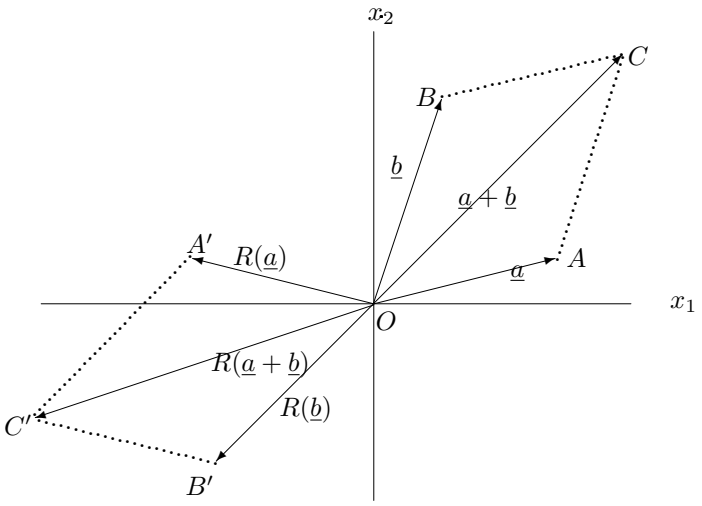


Figure 3.4

We will now see that $R : \Re^2 \to \Re^2$ is a linear transformation.

In Figure 3.5, A and B are any two points in the plane with position vectors $\underline{a}$ and $\underline{b}$. The point C has position vector $\underline{c} = \underline{a} + \underline{b}$ and the quadrilateral $OACB$ is a parallelogram.

The points A', B' and C' are the respective results of rotating about O the points A,B and C through the angle θ. In other words, A', B' and C' have position vectors $R(\underline{a})$, $R(\underline{b})$ and $R(\underline{a}+\underline{b})$ respectively. The parallelogram $OA'C'B'$ is congruent to $OACB$ and this shows that

$$R(\underline{a}+\underline{b}) = R(\underline{a}) + R(\underline{b})$$

Thus the transformation R satisfies the first condition in equation (3.35) for linearity. It is pretty obvious that if we multiply $\underline{a}$ by a scalar s that $R(\underline{a})$ will be multiplied by the same factor, i.e. that $R(\underline{a}s) = R(\underline{a})\,s$, so that that the second condition for linearity as expressed in equation (3.36) also holds. We have shown that $R : \Re^2 \to \Re^2$ is a linear transformation.

3.3.12 The matrix R_θ of a rotation through θ

Let R_θ be the matrix of the above rotation transformation R that rotates through the angle θ. Then

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.53}$$

To see this, note that from Figure 3.5,

$$R_\theta \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad \text{and} \quad R_\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

Since these are the first and second columns of R_θ respectively, this proves (3.53).

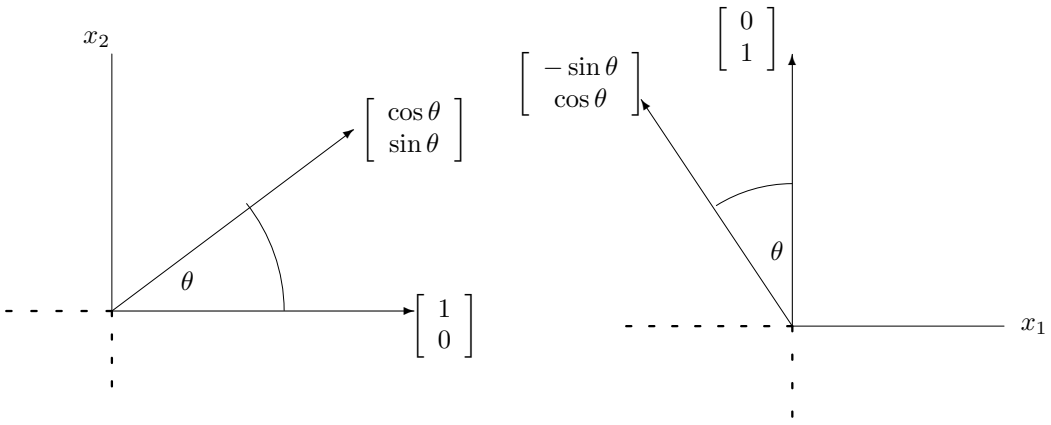


Figure 3.5

Example 3.3.10 *Write down the linear transformations that rotate $\pi/3$ and $2\pi/3$ radians about the origin in the Cartesian plane.*

Solution:

$$R_{\frac{\pi}{3}} = \begin{bmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2} \end{bmatrix}$$

$$R_{\frac{2\pi}{3}} = \begin{bmatrix} \cos\frac{2\pi}{3} & -\sin\frac{2\pi}{3} \\ \sin\frac{2\pi}{3} & \cos\frac{2\pi}{3} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix}$$

3.3.13 The addition formulae for sine and cosine

We note that rotation through ϕ followed by the rotation through θ is equivalent to the rotation through $\theta + \phi$, that is,

$$R_\theta R_\phi = R_{\theta+\phi}$$

This means

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{bmatrix} \tag{3.54}$$

Multiplying the two matrices on the left gives the addition formulae for sine and cosine:

$$\begin{bmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{bmatrix} = \begin{bmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{bmatrix}$$

Remark 62 *Let $X \neq O$ be a point in three-dimensional space. Looking along $\underline{OX}$ we can rotate clockwise through θ and this defines a transformation $T : \Re^3 \to \Re^3$. Much the same argument shows that T is linear. However, the formula for its matrix is more complicated.*

Exercise 63 *1. Find all possible products of the following matrices, whenever the product is defined,and when this is the case, express the rows and columns of the product as linear combinations, as in example 3.3.7.*

$$A = \begin{bmatrix} -1 & 3 \\ 2 & 1 \end{bmatrix},\ B = \begin{bmatrix} -2 & 5 \\ 4 & -7 \\ 0 & 3 \end{bmatrix},\ C = \begin{bmatrix} 2 & 0 & -5 \\ -4 & 3 & 11 \end{bmatrix},$$
$$D = \begin{bmatrix} -3 & 4 & -1 \end{bmatrix},\quad E = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}$$

2. *Given that B is a 3×3 matrix, find by inspection matrices A and C such that*

$$AB = \begin{bmatrix} B_{1\bullet} - 2B_{3\bullet} \\ 2B_{2\bullet} + 5B_{3\bullet} \\ B_{1\bullet} + 4B_{2\bullet} + 7B_{3\bullet} \\ -3B_{1\bullet} \end{bmatrix}$$

$$BC = \begin{bmatrix} B_{\bullet 1}(-2) + B_{\bullet 2}(3) + B_{\bullet 3} & B_{\bullet 2} - B_{\bullet 1} \end{bmatrix}$$

Note that BC has two columns.

3. *If $\underline{a} \in \Re^{1\times n}$ is a row and $\underline{b} \in \Re^{m\times 1}$ is a column, is the product $X = \underline{b}\,\underline{a}$ defined? If so, describe the entries $[X]_{ij}$.*

4. *If $AB = C$ and two of A ,B, C are square then so is the other. True or false?*

5. *Let A be an $m \times n$ matrix and I_m and I_n identity matrices (see equation (3.8)). Show that*

$$I_m A = A = A I_n$$

Conclude that if A is $n \times n$ and $I = I_n$ that $IA = AI = A$.

6. *In No.10 of Exercise 5 of Chapter 1 you found the formula for the projection of a point (p_1, p_2, p_3) on the line $(-2t, -t, 5t)$. Find its matrix B as a linear transformation. In No.1 of Exercise 56 above you found the matrix A of the projection on the plane $2x + z = 0$. Find*

   (a) *the matrix of the transformation that first projects on the plane $2x + z = 0$ then on the line $(-2t, -t, 5t)$.*

   (b) *the matrix of the transformation that first projects on the line $(-2t, -t, 5t)$ then on the plane $2x + z = 0$.*

   The answer is

$$AB = \frac{1}{30}\begin{bmatrix} \frac{1}{5} & 0 & -\frac{2}{5} \\ 0 & 1 & 0 \\ -\frac{2}{5} & 0 & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 4 & 2 & -10 \\ 2 & 1 & -5 \\ -10 & -5 & 25 \end{bmatrix} = \begin{bmatrix} \frac{4}{25} & \frac{2}{25} & -\frac{2}{5} \\ \frac{1}{15} & \frac{1}{30} & -\frac{1}{6} \\ -\frac{8}{25} & -\frac{4}{25} & \frac{4}{5} \end{bmatrix}$$

7. (a) *complete the proof of the addition formulae for sine and cosine by multiplying out equation (3.54).*

   (b) *Check by direct multiplication that $R_\theta R_{-\theta} = I_2$.*

8. ***

   (a) *Let ℓ be a line in the Cartesian plane making an angle θ with the positive x_1-axis. Find the projection matrix P_θ on ℓ and deduce that the matrix M_θ which reflects in ℓ is given by*

$$M_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$$

   Observe that, in spite of superficial appearances, M_θ is not a rotation matrix. Why?

   (b) *Let ϕ and θ be two angles and $\alpha = \phi - \theta$. Prove that $M_\phi M_\theta = R_{2\alpha}$ in two ways:*

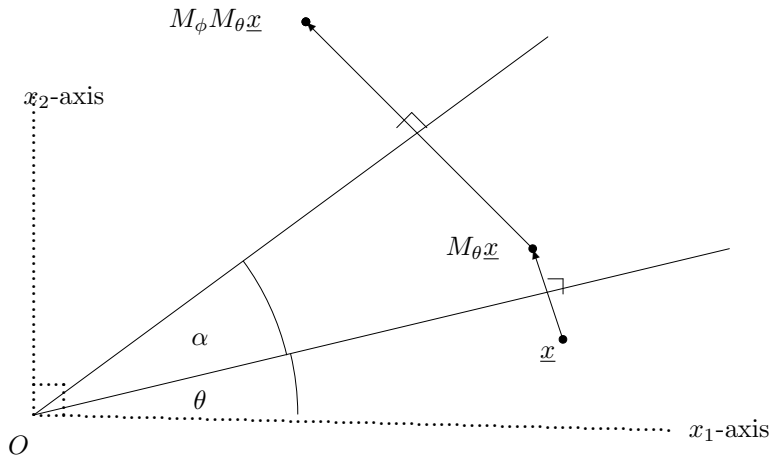


Figure 3.6 $\alpha = \phi - \theta$

i. geometrically by referring to Figure 3.6;

ii. algebraically.

Hint 64 *For the geometric solution, let M_θ rotate $\underline{x}$ through 2γ. Then M_ϕ rotates $M_\theta \underline{x}$ through $2(\alpha - \gamma)$. For the algebraic solution use the expression for M_θ found in the previous exercise.*

9. *Following the corkscrew rule, find matrices that represent rotations through θ about the $x-$, $y-$ and $z-$ axes.*

   Hint 65 *Recall the corkscrew rule: looking along the specified axis, rotate clockwise through θ. The required rotation about the $z-$axis is*

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

10. *(Compare Exercise 56, No.6 above). Let M be the matrix of a reflection in a plane going through the origin and P the matrix of a projection on the same plane. Explain geometrically why you expect the following to be true. Draw diagrams!*

    (a) $M^2 = I_3$

    (b) $P^2 = P$

    (c) $MP = P = PM$

    Verify these results for the plane $2x + z = 0$ of No.1 of Exercise 56 above.
    For analytical proofs of these results see No.17 below

11. *If the matrix product AB is defined, show that so is $B^T A^T$ and that $(AB)^T = B^T A^T$.*

    Hint 66 $[(AB)^T]_{ji} = [AB]_{ij} = A_{i\bullet} B_{\bullet j}$. *Now use equation (3.13) and Exercise 47 No.6.*

12. *Prove that the composite transformation $\underline{x} \to A(B\underline{x})$ of (3.41) is indeed linear. Can you see why a similar result holds for general linear transformations?*

Hint 67 *Consider* $A[B(\underline{u}s + \underline{v}t)] = A[(B\underline{u})s + (B\underline{v})t] = \cdots$

$A[B(\underline{u}s + \underline{v}t)] = A[(B\underline{u})s + (B\underline{v})t] = [A(B\underline{u})]s + [A(B\underline{v})]t$

13. *("Linear combinations of linear combinations" theorem)*

    (a) *In Exercise 47, No.8c you anticipated the following result: Let $\underline{x}$ be a column vector with m entries that depends linearly on $\underline{a}_1 \ldots, \underline{a}_k$ and suppose that each vector $\underline{a}_j$ depends linearly on $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$. Then $\underline{x}$ depends linearly on $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$. Using matrix multiplication give a short proof of this statement.*
    It should be clear that the same result must be true if the vectors are rows, but give a modified matrix proof.

    (b) *Interpret this result in terms of the span concept.*

Hint 68 *Let the $m \times k$ matrix A have columns $\underline{a}_1, \ldots, \underline{a}_k$ and let the $m \times n$ matrix B have columns $\underline{b}_1, \ldots, \underline{b}_n$. By assumption, there is an $n \times k$ matrix C such that $A = BC$. Since $\underline{x} = A\underline{u}$ for some column vector $\underline{u}$ with k entries, the result follows. Fill in the remaining details.*
For the row case, think in terms of the transposes of the above matrices.

14. *Show that the column case of No.13 can also be proved using $\sum$- notation as follows: Write $\underline{x} = \sum_{j=1}^{k} \underline{a}_j u_j$ and $\underline{a}_j = \sum_{i=1}^{n} \underline{b}_i c_{ij}$ for $j = 1, \ldots, n$. Then expand and rearrange the sum*

$$\underline{x} = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} \underline{b}_i c_{ij} \right) z_j$$

15. *Prove that the composite transformation $\underline{x} \to A(B\underline{x})$ of (3.41) is indeed linear.*
Can you see why a similar result holds for general linear transformations?

Hint 69 *Consider* $A[B(\underline{u}s + \underline{v}t)] = A[(B\underline{u})s + (B\underline{v})t] = \cdots$

$A[B(\underline{u}s + \underline{v}t)] = A[(B\underline{u})s + (B\underline{v})t] = [A(B\underline{u})]s + [A(B\underline{v})]t$

16. (a) *If the matrix sum $A + B$ and the matrix product $C(A + B)$ are defined, show that $C(A + B) = CA + CB$.*
    (b) *If the matrix sum $A + B$ and and the matrix product $(A + B)C$ are defined, show that $(A + B)C = AC + BC$.*
    (c) *If AB is defined, and s is a scalar, show $s(AB) = (sA)B$ and $(AB)s = A(Bs)$ and that both are equal. Conclude with statements and proofs of more general results than (a) and (b).*

Hint 70 *To show (a) use the definition of matrix multiplication and equation (3.14):*

$$[C(A + B)]_{ij} = C_{i\bullet}[A + B]_{\bullet j} = C_{i\bullet}(A_{\bullet j} + B_{\bullet j}) = C_{i\bullet}A_{\bullet j} + C_{i\bullet}B_{\bullet j} = \cdots$$

If C is $m \times k$ and A and B are $k \times n$, you may prefer to proceed directly:

$$[C(A + B)]_{ij} = \sum_{r=1}^{k} c_{ir}[A + B]_{rj} = \sum_{r=1}^{k} c_{ir}(a_{rj} + b_{rj}) = ...$$

*17. **

(a) *Let C be an $n \times n$ matrix satisfying $C^2 = I$ where $I = I_n$.*

Show that $(I - C)^2 = I - C$ and that $(2C - I)^2 = I$.

The rest of the question refers to No.10 above and to Exercise 56, No.2a, No.2b and No.4.

(b) *Let P be the projection matrix of Exercise 56, No.2a and write the vector $\underline{u}$ as a column vector $\underline{u} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T$.*

Show that

i. $P = \frac{1}{|\underline{u}|^2} \underline{u}\, \underline{u}^T$,

ii. $P^2 = P$ using matrix multiplication,

iii. the matrix $M = 2P - I$ reflecting in the line satisfies $M^2 = I$.

(c) *Let P be the projection matrix of Exercise 56, No.2b onto a plane through the origin with normal $\underline{n} = \begin{bmatrix} n_1 & n_2 & n_3 \end{bmatrix}^T$.*

Show that

i. $P = I - \frac{1}{|\underline{n}|^2} \underline{n}\, \underline{n}^T$,

ii. $P^2 = P$ using matrix multiplication,

iii. the matrix $M = 2P - I$ reflecting in the plane satisfies $M^2 = I$.

(d) *Complete analytical proofs of No.10 above.*

Hint 71 *Note that we are forming matrix products of a column and a row as in No.3. Also make use of No.17a.*

3.4 Linear independence

In $\Re^2$ let $\underline{a}$ and $\underline{b}$ be two column vectors. If if one depends linearly on the other, say $\underline{a} = \underline{b}s$ for a scalar s, we say they are linearly dependent. Otherwise the vectors are linearly independent: neither depends on the other. In that case, if some linear combination $\underline{a}s + \underline{b}t = \underline{0}$, then both s and t have to be zero. For if (say) $t \neq 0$, then $\underline{b} = \underline{a}\left(-\frac{s}{t}\right)$, and $\underline{b}$ would depend on $\underline{a}$.
Similarly, in $\Re^3$ three column vectors $\underline{a}$, $\underline{b}$ and $\underline{c}$ are linearly dependent if one of them is a linear combination of the other two. This will be the case if there are numbers s, t, r **not all zero** such that $\underline{a}s + \underline{b}t + \underline{c}t = \underline{0}$. For example, if $s \neq 0$, then $\underline{a}$ will depend linearly on $\underline{b}$ and $\underline{c}$. They are linearly independent if no vector depends linearly on the other two. In that case we can only have $\underline{a}s + \underline{b}t + \underline{c}r = \underline{0}$ if $s = t = r = 0$.

Geometrically let $\underline{a}$, $\underline{b}$ and $\underline{c}$ be linearly independent and let the tails of these vectors are placed at the origin O. Then O and the heads of $\underline{a}$, $\underline{b}$ and $\underline{c}$ *do not lie on a plane*. For more details on the geometric meaning of linear independence see subsection 3.4.11 below.
We now generalize and formalize these ideas as follows:

3.4.1 Column independence and column dependence

Definition

Let $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$ be a list of n column vectors each with m entries. Form the $m \times n$ matrix A with these vectors as columns:

$$A = \begin{bmatrix} \underline{a}_1 & \underline{a}_2 & \cdots & \underline{a}_n \end{bmatrix}$$

If there is a **non-trivial** solution $\underline{x} \neq \underline{0}$ to the homogeneous system

$$A\underline{x} = \underline{0} \tag{3.55}$$

the vectors $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$ are said to be **linearly dependent**, otherwise they are **linearly independent**.

To put it another way, the columns of the $m \times n$ matrix A are linearly independent if, and only if for all column vectors $\underline{x}$ with n entries

$$A\underline{x} = \underline{0} \quad \text{implies} \quad \underline{x} = \underline{0} \tag{3.56}$$

To say the same thing once more, if we can find a linear combination such that

$$\underline{a}_1 x_1 + \underline{a}_2 x_2 + \cdots + \underline{a}_n x_n = \underline{0} \tag{3.57}$$

in which **one or more** of the coefficients x_j are not zero, then the vectors $\underline{a}_1$, $\underline{a}_2$,..., $\underline{a}_n$ are linearly dependent (there is a non-trivial linear relationship among them), otherwise they are independent.

3.4.2 An equivalent way to see independence/dependence

Let $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$ be column vectors each with m entries.
If there is only one vector, *i.e.* $n = 1$, then $\underline{a}_1$ is dependent if, and only if it is the zero vector.
If $n \geq 2$, the vectors are linearly dependent if, and only if, one of them depends linearly on the others.

Proof:
The proof of the case $n = 1$ is left to you. See Exercise 78, No.1a.
Let $n \geq 2$ and suppose the vectors are linearly dependent. Then (say) $x_1 \neq 0$ in equation 3.57.
In that case $\underline{a}_1$ depends linearly on $\underline{a}_2$,..., $\underline{a}_n$:

$$\underline{a}_1 = \underline{a}_2 \left(-\frac{x_2}{x_1}\right) + \cdots + \underline{a}_n \left(-\frac{x_n}{x_1}\right)$$

Similarly, if $x_j \neq 0$, then $\underline{a}_j$ depends linearly on the vectors $\underline{a}_i$ for $i \neq j$.
Conversely, if (say) there are scalars $s_2, \ldots, s_n$ such that

$$\underline{a}_1 = \underline{a}_2 s_2 + \cdots + \underline{a}_n s_n$$

Then

$$\underline{a}_1 + \underline{a}_2 (-s_2) + \cdots + \underline{a}_n (-s_n) = \underline{0}$$

Since the coefficient of $\underline{a}_1$ is $1 \neq 0$, the vectors are linearly dependent, as required.

For a slightly refined version of this theorem, see Exercise 78, No.1(c)v.

Independence/dependence of row vectors

The definition of linear independence for a list of row vectors is almost identical to that for column vectors. See 3.4.12 below.

3.4.3 Some examples of independence/dependence

The most elementary example of linear independence

The simplest example of linearly independent column vectors is afforded by one or more columns of the $n \times n$ identity matrix $I = I_n$. Since $I\underline{x} = \underline{x}$, we can only have $I\underline{x} = \underline{0}$ if $\underline{x} = \underline{0}$. Slightly more complicated is

Example 3.4.1 *Consider the matrix (3.25) of example 3.1.6:*

$$A = \begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix}$$

A row-echelon was found in equation (3.26):

$$A' = \begin{bmatrix} 2 & 5 & -11 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

The columns of A are linearly dependent, as can be seen from (3.27) by letting $t \neq 0$, say putting $t = 1$.

*The columns of the matrix (3.29) in example 3.1.7 are linearly independent since $A\underline{x} = \underline{0}$ has **only** the trivial solution $\underline{x} = \underline{0}$.*

The columns of A in Exercise 35 No.2b are linearly independent.

3.4.4 If a matrix has more columns than rows, its columns are linearly dependent; if more rows than columns then its rows are linearly dependent

This is just a rephrasing of the result of subsection 2.2.15 in Chapter 2: A homogeneous system with more unknowns than equations has a non-trivial solution.

The row case is quite similar.

3.4.5 Testing for linear independence of some columns of A

Suppose that A has seven columns and we wish to test if the columns of $[A_{\bullet 2}, A_{\bullet 5}, A_{\bullet 6}]$ are linearly independent. This is the same as considering all vectors $\underline{x}$ with $x_1 = x_3 = x_4 = 0$ and testing the implication (3.56). The same principle obviously applies to any selection of columns of A.

3.4.6 Column-rank and row-rank

If $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ are column vectors, their **column-rank** is the **maximum number** of vectors $\underline{a}_j$ that are linearly independent. This is also the column-rank of the matrix $A = \begin{bmatrix} \underline{a}_1 & \underline{a}_2 & \cdots & \underline{a}_n \end{bmatrix}$.

Row-rank The row-rank of a matrix A is defined similarly, except that 'rows' replace 'columns'. See 3.4.12 below for more details.

Given a list of vectors of the same size (either rows or columns), it should be clear what is be meant by their **rank**: it is the maximum number of linearly independent vectors among them.

3.4.7 Elementary row operations preserve column independence and dependence

Consider a system $A\underline{x} = \underline{b}$ of m equations for n unknowns. Let $\left[A', \underline{b}'\right]$ be the array after performing elementary row operations. We saw in the last chapter that $A\underline{x} = \underline{b}$ holds if, and only if, $A'\underline{x} = \underline{b}'$ holds:

$$\underline{b} = A_{\bullet 1}x_1 + \cdots + A_{\bullet n}x_n \quad \text{if, and only if,} \quad \underline{b}' = A'_{\bullet 1}x_1 + \cdots + A'_{\bullet n}x_n$$

Now for the homogeneous system (3.55), $\underline{b} = \underline{b}' = \underline{0}$, and any dependency relation among the columns of A also holds for A' and *vice-versa* :

$$A_{\bullet 1}x_1 + \cdots + A_{\bullet n}x_n = \underline{0} \quad \text{if, and only if,} \quad A'_{\bullet 1}x_1 + \cdots + A'_{\bullet n}x_n = \underline{0}$$

In particular, certain columns of A are linearly independent if, and only if, **the same** columns of A' are linearly independent. In particular, the column ranks of A and A' are the same.
Also, column $A_{\bullet j}$ is a linear combination of other columns of A if, and only if, $A'_{\bullet j}$ is a linear combination of the corresponding columns of A' with the same coefficients.

For these reasons, questions regarding dependence or independence of the columns of A are best answered by considering the **row reduced echelon form** (row REF) of A.

3.4.8 The column-rank of a matrix is the number of non-zero rows in a row echelon form

Proof

Let k be the number of non-zero rows in an echelon form of the $m \times n$ matrix A and suppose that A'' is its row REF. Recall that the pivot columns (see 2.2.8) of A'' are columns of the identity matrix I_m, which are certainly linearly independent. Hence the column-rank of A'' is at least k. The $k \times n$ matrix of the non-zero rows of A'' cannot have column-rank more than k by 3.4.4. Since the other rows of A'' are zero, the column rank of A'' is exactly k. This is also the column-rank of A.

3.4.9 Every column of A is a linear combination of the linearly independent pivot columns of A

This is because the pivot columns of A'', when restricted to the non-zero rows, are the columns of the identity matrix I_k.
The statement says that the pivot columns span the column space.

Example 3.4.2 *Consider example 3.1.6 once more. Can we find a set of linearly independent columns of A in (3.25) such that every column of A is a linear combination of these independent columns? What is the column-rank of A? Illustrate the statements of 3.4.8 and 3.4.9.*

Solution:

The original matrix (3.28) is

$$A = \begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix}$$

The echelon form 3.26 is

$$\begin{bmatrix} 2 & 5 & -11 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

shows that first two columns of A are linearly independent. The row REF shows things even more clearly:

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.58}$$

Every column of A is a linear combination of the first two, which are pivot columns. The column rank is 2.

Example 3.4.3 *Let*

$$A = \begin{bmatrix} -1 & -4 & -3 & 2 \\ 3 & 4 & 1 & 2 \\ 2 & 9 & 7 & -5 \\ -5 & -1 & 4 & -9 \end{bmatrix}$$

Illustrate 3.4.7, 3.4.8 and 3.4.9.

Solution:

We first reduce the matrix A to row REF:

$$\begin{array}{cccc} -1 & -4 & -3 & 2 & \\ 0 & -8 & -8 & 8 & 3R_1 + R_2 \\ 0 & 1 & 1 & -1 & 2R_1 + R_3 \\ 0 & 19 & 19 & -19 & -5R_1 + R_4 \end{array}$$

$$\begin{array}{cccc} -1 & -4 & -3 & 2 & \\ 0 & -1 & -1 & 1 & \frac{1}{8}R_2 \\ 0 & 1 & 1 & -1 & \\ 0 & 1 & 1 & -1 & \frac{1}{19}R_4 \end{array}$$

$$\begin{array}{cccc} -1 & -4 & -3 & 2 & \\ 0 & -1 & -1 & 1 & \\ 0 & 0 & 0 & 0 & R_2 + R_3 \\ 0 & 0 & 0 & 0 & R_2 + R_4 \end{array}$$

$$\begin{array}{cccc} -1 & -4 & -3 & 2 & \\ 0 & 1 & 1 & -1 & -R_2 \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \end{array}$$

$$\begin{array}{cccc} -1 & 0 & 1 & -2 & 4R_2 + R_1 \\ 0 & 1 & 1 & -1 & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \end{array}$$

This gives the row REF A'' of A as

$$E = \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivot columns of E are $E_{\bullet 1}$ and $E_{\bullet 2}$ and they are linearly independent, and therefore so are $A_{\bullet 1}$ and $A_{\bullet 2}$. The column-rank of A is 2. Clearly, every column of E is a linear combination of $E_{\bullet 1}$ and $E_{\bullet 2}$. Therefore, every column of A depends linearly on $A_{\bullet 1}$ and $A_{\bullet 2}$. For example, $E_{\bullet 4} = E_{\bullet 1}(2) + E_{\bullet 2}(-1)$, and so $A_{\bullet 4} = A_{\bullet 1}(2) + A_{\bullet 2}(-1)$, much as in the previous example.

3.4.10 Uniqueness of the solution to $A\underline{x} = \underline{b}$. One-to-one linear transformations

If the columns of the $m \times n$ matrix A are linearly independent then for any $m \times 1$ column vector $\underline{b}$ the system of equations $A\underline{x} = \underline{b}$ has **at most** one solution.

To see this, consider the following:

Suppose that $\underline{u}$ and $\underline{v}$ are two solutions, i.e. $A\underline{u} = \underline{b}$ and $A\underline{v} = \underline{b}$. Then $A\underline{u} = A\underline{v}$ and so $A\underline{u} - A\underline{v} = A(\underline{u} - \underline{v}) = \underline{0}$. As the columns of A are linearly independent, $\underline{u} - \underline{v} = \underline{0}$ and $\underline{u} = \underline{v}$.

Consider the matrix mapping $\underline{x} \to A\underline{x}$ of (3.32). To say that this mapping is **one-to-one** means that for any two column vectors $\underline{u}$ and $\underline{v}$ with n entries, we can have $A\underline{u} = A\underline{v}$ only if $\underline{u} = \underline{v}$.
In other words, two linear combinations of the columns of A are equal only if they have identical coefficients. The mapping $\underline{x} \to A\underline{x}$ is one-to-one if, and only if, the column-rank of A is n.

Example 3.4.4 *A row echelon form (2.2) of the coefficient matrix A of Example 2.1.2 in Chapter 2 is*

$$\begin{bmatrix} 1 & 1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

The column-rank of A is 3 *and the mapping*

$$\underline{x} \to \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 4 \\ -1 & 1 & -2 \end{bmatrix} \underline{x}$$

from $\Re^3$ *to* $\Re^3$ *is one-to-one.*
All this becomes even clearer if we look at the row-reduced echelon form in Example 2.2.7 of Chapter 2. This is the 3×3 *identity matrix and so* $A\underline{x} = \underline{b}$ *has a solution for any* $\underline{b} \in \Re^3$*: The mapping* $\underline{x} \to A\underline{x}$ *has range the whole of* $\Re^3$*.*
The example illustrates the following remark.

3.4.11 Geometric interpretation of column-rank

A proper definition of 'dimension' is given in section 3.6), but we can accept that a line in ordinary space through O has dimension 1, that a plane through O has dimension 2 and that the dimension of the whole of ordinary space is 3. The column-rank of a matrix A can be thought of as the **dimension of its range** when A is considered as a linear mapping. To better appreciate this, see Exercise 78, No.19 as well as No.8 in Exercise 93.
Consider the matrix of Example 3.3.1 that projects on a line. Each column is a multiple of any other and the column-rank is 1, as expected.
Again, consider the coefficient matrix A of equation (3.40). Because $A_{\bullet 1} + A_{\bullet 2}(2) + A_{\bullet 3} = \underline{0}$, the column-rank of A is not 3. It is easy to see that any two columns are linearly independent, so $r(A) = 2$, in agreement with our comment.
The range of a reflection matrix is the whole of space, which has dimension 3. So we expect its rank to be 3, which is true.

3.4.12 Row Independence and row-rank

Everything that can be said about column independence has its counterpart for rows. Hence, as indicated earlier, the **rows** of the $m \times n$ matrix A are linearly independent if, and only if for all row vectors $\underline{y}$ with m entries,

$$\underline{y}A = \underline{0} \quad \text{implies} \quad \underline{y} = \underline{0}$$

In other words, if we can find a row vector $\underline{y} \neq \underline{0}$ such that the linear combination

$$y_1 A_{1\bullet} + y_2 A_{2\bullet} + \cdots + y_m A_{m\bullet} = \underline{0}$$

then the rows of A are linearly dependent, otherwise they are linearly independent.

The **row-rank** of A is the maximum number of linearly independent rows of A.

3.4.13 The non-zero rows of a matrix in row echelon form are linearly independent

To see this, let C be an $m \times n$ matrix in row echelon form, and suppose that $C_{1\bullet}$, $C_{2\bullet}$,...., $C_{k\bullet}$, are its non-zero rows. Let

$$y_1 C_{1\bullet} + y_2 C_{2\bullet} + \cdots + y_k C_{k\bullet} = \underline{0}$$

where $\underline{0}$ is the zero row with n entries. The first non-zero entry in $C_{1\bullet}$ has only zeros below it. Therefore, $y_1 = 0$ and $y_2 C_{2\bullet} + \cdots + y_k C_{k\bullet} = \underline{0}$. As the first non-zero entry in $C_{2\bullet}$ has only zeros below it, $y_2 = 0$. Carrying on the argument, $y_1 = y_2 = \cdots = y_k = 0$. The claim is proved.

For example, the first two rows of

$$\begin{bmatrix} -1 & -4 & -3 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

For more examples see Chapter 2, including Exercises 35 and 39.

3.4.14 Elementary column operations for equations of type $\underline{y}A = \underline{c}$

Let A be an $m \times n$ matrix. In order to solve the matrix equation

$$\underline{y}A = \underline{c} \tag{3.59}$$

we can mimic the procedure followed to solve a system of the type $A\underline{x} = \underline{b}$. To do this we must perform elementary column operations on the augmented array

$$\begin{bmatrix} A \\ \underline{c} \end{bmatrix}$$

Such operations **do not alter the solution set** to equation 3.59 and are:

(i) interchange two columns of the array, (ii) multiply a column of the array by a non-zero constant and (iii) add a multiple of some column to another.
In remark 49 we mentioned that we are in the habit of rather doing elementary row operations on the system

$$A^T \underline{y}^T = \underline{c}^T$$

Every elementary row operation on this system has its exact counterpart as an elementary column operation on the system $\underline{y}A = \underline{c}$, and this is true in particular for $\underline{c} = \underline{0}$.

Nevertheless, column operations on an array are important in their own right.

Example 3.4.5 *Let*

$$B = \begin{bmatrix} -1 & 3 & 2 & -5 \\ -4 & 4 & 9 & -1 \\ -3 & 1 & 7 & 4 \\ 2 & 2 & -5 & -9 \end{bmatrix}$$

Find the row-rank of B.

Solution:

B is the transpose of the matrix A in example 3.4.3. Thus its row-rank is the column-rank of A, which is 2.

Note 72 *For further elementary properties of linear dependence/independence, see No.1 in Exercise 78 below.*

3.4.15 A Column Lemma

Let A and B have the same number n of columns. If for all column vectors $\underline{x}$ with n entries,

$$A\underline{x} = \underline{0} \quad \text{implies} \quad B\underline{x} = \underline{0} \tag{3.60}$$

then (column-rank A) $\geq$ (column-rank B).

Proof: We may suppose that the first k columns of B are independent, where k is the column-rank of B. Let $\underline{x}$ be any vector such that $x_{k+1} = \cdots = x_n = 0$ and $A\underline{x} = \underline{0}$. By assumption, $B\underline{x} = \underline{0}$. As the first k columns of B are independent, $\underline{x} = \underline{0}$. Hence the first k columns of A are independent.

Remark 73 *We have actually shown a bit more: if the condition (3.60 holds, and certain columns of B are linearly independent, then the corresponding columns of A are also linearly independent.*

3.4.16 Corollary: column-rank of a matrix product

Let the matrix product $CA = B$ be defined. Then (column-rank $A \geq$ (column-rank B) and if C has independent columns, the two ranks are equal.

For, if $A\underline{x} = \underline{0}$, then certainly $B\underline{x} = \underline{0}$ and so (column -rank A) $\geq$ (column-rank B) by the Lemma 3.4.15.
Now let C have independent columns and suppose $B\underline{x} = \underline{0}$. Then $C(A\underline{x}) = \underline{0}$ and therefore $A\underline{x} = \underline{0}$ as the columns of C are independent. It follows, again by the Lemma, that (column-rank B) $\geq$ (column-rank A) and equality holds.

Example 3.4.6 *Let C be the 3×3 matrix of Example 3.4.4 of rank 3 and let A be the 3×3 matrix of example 3.4.2 with column-rank 2. From the corollary we can conclude that the column-rank of the product*

$$CA = \begin{bmatrix} 1 & 1 & 3 \\ -1 & 1 & -2 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix} = \begin{bmatrix} 11 & 20 & -38 \\ -13 & -25 & 49 \\ 15 & 27 & -51 \end{bmatrix}$$

is also 2.

The counterpart for rows of the Column Lemma is

3.4.17 A Row Lemma

Let A and B have the same number m of rows. If for all row vectors $\underline{y}$ with m entries,

$$\underline{y}A = \underline{0} \quad \text{implies} \quad \underline{y}B = \underline{0}$$

then row-rank $A \geq$ row-rank B.

Remark 74 *As with the Column Lemma, we have shown that if the condition (3.4.17) holds and certain rows of B are linearly independent, the corresponding rows of A are also linearly independent.*

3.4.18 Corollary: row-rank of a matrix product

Let $AD = B$ be a matrix product. Then row-rank $A \geq$ row-rank B and equality holds if D has independent rows.

The proof of this Lemma and its corollary are left as exercises. See No.13 in Exercise 78.

3.4.19 The the column-rank and row-rank of a matrix A are equal

Proof:

Let the column-rank of the $m \times n$ matrix A be k. By 3.4.8 the k pivot columns of A are linearly independent, forming an $m \times k$ matrix B. By 3.4.9 all columns of A are linear combinations of the columns of B. Thus, for some for some $k \times n$ matrix D,

$$A = BD$$

By Corollary 3.4.18, row-rank $B \geq$ row-rank A. But B has k columns, so its row-rank cannot exceed k:

$$\text{column-rank } A = k \geq \text{row-rank } B \geq \text{row-rank } A$$

Hence the column-rank of A is at least the row-rank of A.

By symmetry (consider the transpose A^T), the row-rank of A is at least the column-rank of A and so the two are equal.

3.4.20 Definition of rank $r(A)$

The common row- and column-rank of A is called the **rank** of A and is denoted by $r(A)$.
By 3.4.8, the rank of A is the number of non-zero rows in a row echelon form of A.

Example 3.4.7 *Find the rank of the matrix*

$$A = \begin{bmatrix} 0 & 0 & -3 & -3 & -1 \\ 2 & -2 & -2 & -3 & -2 \\ 0 & 0 & 3 & -2 & -2 \\ 2 & -3 & -4 & 3 & -3 \end{bmatrix}$$

Solution:

This is the matrix of Example 2.2.4 in Chapter 2. We found its row-echelon form in equation 2.26:

$$\begin{bmatrix} 2 & -3 & -4 & 3 & -3 \\ 0 & 1 & 2 & -6 & 1 \\ 0 & 0 & 3 & -2 & -2 \\ 0 & 0 & 0 & -5 & -3 \end{bmatrix}$$

Therefore, $r(A) = 4$

3.4.21 Lemma: A basic property of linear independence

A basic property concerning linear independence/dependence is the following.

Let $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ be a list of n linearly independent column vectors, each with m entries and $\underline{b}$ a column vector with m entries. Then the rank of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, $\underline{b}$ equals the rank of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ if, and only if, $\underline{b}$ depends linearly on $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$.

Proof:
The rank of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b}$ is either n or $n+1$. If it is n, then these vectors are linearly dependent, and there exist constants x_1,...,x_n, β **not all zero** such that

$$\underline{a}_1 x_1 + \underline{a}_2 x_2 + \cdots + \underline{a}_n x_n + \underline{b}\beta = \underline{0}$$

We must have $\beta \neq 0$, otherwise one or more of the scalars x_j would have to be non-zero, contradicting the linear independence of the vectors $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$. Therefore,

$$\underline{b} = \underline{a}_1 \left(-x_1\beta^{-1}\right) + \underline{a}_2 \left(-x_2\beta^{-1}\right) + \cdots + \underline{a}_n \left(-x_n\beta^{-1}\right)$$

and $\underline{b}$ is a linear combination of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$.
Conversely, if $\underline{b}$ is a linear combination of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, then $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b}$ cannot be linearly independent, and Theorem A has been proved.

The Lemma rephrased: Given linearly independent vectors $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, the vectors $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b}$ are linearly independent if, and only if, $\underline{b}$ does not depend linearly on $\underline{a}_1, \underline{a}_2, \ldots \underline{a}_n$.

3.4.22 Theorem: The fundamental property of linear independence

Note 75 *This theorem is a generalization of the Lemma. To understand what follows you should have a thorough grasp of the content of the "linear combinations of linear combinations theorem" contained in No.8c of Exercise 47 and No.13 of Exercise 63.*

Let $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, $\underline{b}$ be any list of n column vectors, each with m entries. Then the rank of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, $\underline{b}$ is not more than the rank of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ if, and only if, $\underline{b}$ depends linearly on $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$.

Put slightly differently, the theorem says that

$$\text{column-rank}\,[A, \underline{b}] \;=\; \text{column-rank}\,A$$

if, and only if, $A\underline{x} = \underline{b}$ has a solution. Here, A is the $m \times n$ matrix having the vectors $\underline{a}_j$ as columns.

Proof:

Let the maximum number of linearly independent vectors among $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ be s.

1. Suppose that the rank of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b}$ is s. Then, by Theorem A, $\underline{b}$ must be a linear combination of any s linearly independent vectors among $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$.

2. Conversely (and this is the only non-trivial case), suppose that the rank of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, $\underline{b}$ is $s+1$.
   Then we may assume (without loss in generality) that $\underline{a}_1, \ldots, \underline{a}_s, \underline{b}$ are linearly independent. By Theorem A, every vector $\underline{a}_j$ is a linear combination of $\underline{a}_1, \ldots, \underline{a}_s$. Then if $\underline{b}$ were a linear combination of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, it would also be a linear combination of $\underline{a}_1, \cdots, \underline{a}_s$, which is a contradiction, because $\underline{a}_1, \ldots, \underline{a}_s, \underline{b}$ are assumed to be linearly independent.

Corollary 76 *Let the matrix B consist of k linearly independent columns of a matrix A which cannot be enlarged by columns of A without destroying independence. Then*

1. *Every column of A is a linear combination of columns of B (this follows from the Lemma)*

2. *k is the column-rank of A.*

Remark 77 *We emphasize that the above theorem 3.4.22 about column vectors is equally valid for row vectors. The statement 3.4.22 goes through word-for-word.*
The full use of the theorem will be seen in section 3.6 when we study vector spaces.

3.4.23 An Application of the Fundamental Theorem

We will use 3.4.22 to show directly that elementary row operations on an $m \times n$ matrix A do not change its row-rank.

It is clear that interchanging two rows cannot change the row-rank. It is almost as obvious that multiplication of a row by a non-zero constant does not alter the row-rank (see No.1(c)iv in Exercise 78). It remains to show that replacing (say) $A_{1\bullet}$ by $\beta A_{2\bullet} + A_{1\bullet}$ to form A' does not change the row-rank.
Let the last $m-1$ rows $A_{2\bullet}, \cdots, A_{m\bullet}$ of A have rank k. Since $\beta A_{2\bullet} + A_{1\bullet}$ is a linear combination of these $m-1$ rows if, and only if, $A_{1\bullet}$ is (see No.2 in Exercise 52), it follows that the row-rank of A' is $k+1$ if,and only if, the row-rank of A is $k+1$.

For an alternative way to view this see Exercise 93 No. 9.

Exercise 78 *1. In most of the elementary questions that follow, you may assume the given vectors are either rows or columns, each with the same number of entries. Prove the given statements.*

(a) A single vector $\underline{a}$ is linearly independent if, and only if, $\underline{a} \neq \underline{0}$.

(b) Show directly that two vectors $\underline{a}$, $\underline{b}$ are linearly dependent if, and only if, one is a multiple of the other. Compare the idea of parallel vectors in subsection 1.1.10, Chapter 1.

(c) Let $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$ be a given list of vectors. Show

i. If some $\underline{a}_i = \underline{0}$, then $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$ are linearly dependent.

ii. If $n \geq 2$ and some subset, say $\underline{a}_2, \cdots, \underline{a}_m$, are linearly dependent, so are $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$

iii. If $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$ are linearly independent, what can be said of any subset of $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_m$?

iv. The rank of $\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_n$ is unaltered if one of the vectors $\underline{a}_j$ is multiplied by a non-zero scalar k.

*v. * Let $n \geqslant 2$ and suppose that $\underline{a}_1$, $\underline{a}_2$, ..., $\underline{a}_n$ are **non-zero** vectors. Prove that the vectors are linearly dependent if and only if some $\underline{a}_j$ $(j \geqslant 2)$ depends linearly on $\underline{a}_1$,..., $\underline{a}_{j-1}$. Translate this statement into one about linear independence.*

***Hint 79** Suppose the vectors $\underline{a}_1$, $\underline{a}_2$, ..$\underline{a}_n$ are linearly dependent. Since $\underline{a}_1 \neq \underline{0}$, there must be a first index $j \geqslant 2$ such that the vectors $\underline{a}_1$,..., $\underline{a}_j$ are linearly dependent.*

2. How does Lemma 3.4.21 show why 3.4.9 follows directly from 3.4.8?

3. Let $\underline{a}$, $\underline{b}$, $\underline{c}$ be the position vectors of the points A, B and C respectively. Show that the three vectors are linearly independent if, and only if, A, B, C and O do not lie on a plane. (This was mentioned in the preamble 3.4).

4. Let $A = [C, D]$ where C and D are matrices with the same number of rows. Show that if the rows of C (or of D) are independent, so are those of A. Conclude that in any case $r(A) \geqslant r(C)$ and $r(A) \geqslant r(D)$.

5. For any matrix product AB, say why $r(AB) \leq r(A)$ and $r(AB) \leq r(B)$.

*6. let C be $m \times n$. If A is $m \times m$ of rank m, then $r(AC) = r(C)$. Let B be $n \times n$ of rank n, then $r(CB) = r(C)$. Consequently, $r(ACB) = r(C)$.
Prove these statements.*

7. *Without multiplying out, find the rank of*

$$\begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 \\ -1 & 1 & -2 \\ 1 & 1 & 4 \end{bmatrix} = \begin{bmatrix} -14 & -4 & -48 \\ 15 & 3 & 51 \\ -16 & -2 & -54 \end{bmatrix}$$

Hint 80 *Consider Example 3.4.6.*

8. *The following questions refer you to the matrices in Exercise 35 of Chapter 2. To answer them you should consult the row reduced echelon forms (row REF) found in Exercise 39. Occasionally you may need to find the new row reduced echelon forms.*

   (a) *Consider the matrix A of No.1a.Show that the first two columns of A form a maximum independent set of columns of A and that every column of A is a linear combination of $A_{\bullet 1}$ and $A_{\bullet 2}$. Can you find three linearly independent columns of A? Find the rank $r(A)$.*
   Show that the first two rows of A form a maximum number of independent rows of A.

   Hint 81 *Either show directly that $A_{1\bullet}$ and $A_{2\bullet}$ are linearly independent or use the matrix B of No.1b in Exercise 35.*

   (b) *Let C be the row REF of matrix A of No.2 of Exercise 35.*
   Show that $A_{\bullet 5}$ depends linearly on $A_{\bullet 1}$, $A_{\bullet 2}$, $A_{\bullet 3}$ and that every maximum set of linearly independent columns must include $A_{\bullet 4}$. Find a maximum set of linearly independent columns of A and the rank $r(A)$.

   Which of the following sets are linearly independent?

   i. $A_{\bullet 2}$, $A_{\bullet 3}$, $A_{\bullet 4}$, $A_{\bullet 5}$
   ii. $A_{\bullet 1}$, $A_{\bullet 3}$, $A_{\bullet 4}$, $A_{\bullet 5}$
   iii. $A_{\bullet 1}$, $A_{\bullet 2}$, $A_{\bullet 4}$, $A_{\bullet 5}$

   Find a maximum number of linearly independent solutions to the homogeneous equation.

   (c) *Consider the coefficient matrix A in No.3 of Exercise 35.*

      i. Does any row vector $\underline{c} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix}$ depend linearly on the rows of A ?

      Solution*:*

      A is 4×4 and the rows of A together with $\underline{c}$ can have rank at most four. Since the rows of A are linearly independent ($r(A) = 4$ from the row REF), the answer is 'yes' by the row version of Lemma 3.4.21.

      ii. Are the columns of the coefficient matrix A linearly independent?

      iii. Does $A\underline{x} = \underline{b}$ have a unique solution for every column vector $\underline{b}$ with 4 entries?

      iv. Let C be the augmented matrix of No.3a in Exercise 35. Does every column vector $\underline{b}$ with 4 entries depend linearly on $C_{\bullet 1}$, $C_{\bullet 2}$, $C_{\bullet 3}$, $C_{\bullet 5}$?

9. *Prove that in 3-space mutually orthogonal vectors are linearly independent. Hence complete a rigorous proof of No.15 of Exercise 2 in Chapter 1.*

10. ** Give rigorous proofs of the statements about planes in subsection 1.3.6.*

11. *Let I be the $m \times m$ identity matrix. The matrix J resulting from I after performing an elementary row operation on I is called an $m \times m$* ***elementary matrix****. Hence J is the result of doing one of the following: (i) multiplying row r of I by $\beta \neq 0$, (ii) interchanging rows r and s of I, and (iii) changing row $I_{s\bullet}$ to $\beta I_{r\bullet} + I_{s\bullet}$ if $s \neq r$.*

    (a) *Show that an elementary matrix J has rank m.*

    (b) *Show furthermore that if A is an $m \times n$ matrix, then JA is the result of doing the corresponding elementary row operation on A.*

    Hint 82 *Use $[JA]_{i\bullet} = J_{i\bullet}A$. So, for example, let J be the result of a type (iii) operation. Then for $i \neq s$, $[JA]_{i\bullet} = A_{i\bullet}$, while if $i = s$,*

$$[JA]_{s\bullet} = J_{s\bullet}A = (\beta I_{r\bullet} + I_{s\bullet})A = (\beta I_{r\bullet})A + I_{s\bullet}A = \beta A_{r\bullet} + A_{s\bullet}$$

    (c) *What are the corresponding results for elementary column operations?*

    (d) *Give a description of "column echelon form" and "column reduced echelon form" of a matrix A.*

12. *Show that A, A^T, A^TA and AA^T all have the same rank.*

    Hint 83 *First of all, recall that if $\underline{a}$ is a column vector and $\underline{a}^T\underline{a} = 0$, then $\underline{a} = \underline{0}$ (See Exercise 47 No.11).*
    Now suppose $A^TA\underline{x} = \underline{0}$. Then $\underline{x}^TA^TA\underline{x} = (A\underline{x})^TA\underline{x} = 0$.

13. *Prove the Lemma 3.4.17 and its corollary 3.4.18.*

    Hint 84 *One way: Use A^T and B^T in place of A and B in the previous Lemma 3.4.15. For a more pleasant direct proof, we may suppose that the first $k = r(B)$ rows of B are independent. Let $\underline{y}$ be any row vector such that $y_{k+1} = \cdots = y_m = 0$ and $\underline{y}A = \underline{0}$. Now proceed as in the Column Lemma 3.4.15, except use rows in place of columns.*

    Solution*:*

    By assumption, $\underline{y}B = \underline{0}$. As the first $k = r(B)$ rows of B are independent, $\underline{y} = \underline{0}$. Hence the first k rows of A are independent.

14. *Complete the following alternative proof of the corollary to the Column Lemma (Corollary 3.4.16) which says $r(BC) \leq r(C)$. If k columns (say $[BC]_{\bullet 1}, \cdots, [BC]_{\bullet k}$) of BC are linearly independent then so are $C_{\bullet 1}, \cdots, C_{\bullet k}$. Do the same for the row version.*

15. *In the following assume the relevant matrix products are defined. Prove the statements.*

    (a) *If the columns of A are independent and $AB = AC$, then $B = C$*

    (b) *If the rows of A are independent and $BA = CA$ then $B = C$.*

16. *Let A be $n \times n$ and $A \neq I$, $A \neq O$ and $A^2 = I$, where $I = I_n$. Find all integers $k \geq 1$ such that $A^k = I$.*

17. ** Using (among other things) the Column Lemma 3.4.15, show that if the condition (3.60) holds then $B = CA$ for some C. (This is essentially the converse of the corollary 3.4.16). State and prove the corresponding result for the Row Lemma 3.4.17.*

    Hint 85 *Let $\underline{d} = \underline{d} = B_{i\bullet}$ be a row of B and let A' be the augmented matrix formed by adding row $\underline{d}$ to A. Then $A'\underline{x} = \underline{0}$ if, and only if, $A\underline{x} = \underline{0}$. Hence A' and A have the same column ranks and so the same row ranks. By the row version of the basic theorem 3.4.22 on linear independence, $\underline{d}$ is a linear combination of the rows of A.*

State and prove a similar result corresponding to the Row Lemma 3.4.17.

18. *Find the ranks of the following matrices and prove your results correct.*

    (a) *A is the projection matrix on the plane $2x + z = 0$ of Exercise 56, No.1.*

    (b) *A is the reflection matrix in the plane $2x + z = 0$.*

19. *Let A be a 3×3 matrix. In each of the following cases A is considered as a linear transformation. Describe the rank of A with a brief geometrical proof. Illustrate the comment 3.4.11.*

    (a) *A is a projection matrix onto a line that passes through the origin.*

    (b) *A is a projection matrix onto a plane that passes through the origin.*

    (c) *A is a matrix that reflects in a plane or line that passes through the origin. This again illustrates the remarks in 3.4.11.*

20. ** Use the results of Exercise 63, No.17 to give analytic proofs for No.19 above.*

21. *If R_θ is the 2×2 rotation matrix of equation (3.53), prove algebraically that $r(R_\theta) = 2$.*

22. *Let A be an $m \times n$ matrix.*

    (a) *What is the row REF A'' of A if $r(A) = n$? What is A'' if $r(A) = m = n$?*

    (b) *Let $A\underline{x} = \underline{b}$ be a system of m equations for n unknowns. If the rows of A are independent then the system always has a solution. Find a quick proof using the Fundamental Theorem 3.4.22.*

    (c) *Let $\underline{y}A = \underline{c}$ be a system of n equations for m unknowns $y_1, \cdots, y_m$. If the columns of A are independent then the system always has a solution.*

    (d) *What can be said of the solutions in 22b and 22c if $m = n$?*

3.5 Square matrices

In this section we consider square matrices A of size $n \times n$. Unless otherwise stated, I stands for the identity matrix I_n. Observe that $r(I) = n$.

3.5.1 Invertible matrices

3.5.2 Definition of invertibility

An $n \times n$ matrix A is said to be **invertible** (or **non-singular** or to have an **inverse**) if there is a matrix B such that

$$AB = BA = I \tag{3.61}$$

It is clear that such a matrix B must also be $n \times n$. From the definition it is also clear that if B is an inverse of A then A is an inverse of B.
From Exercise 78 No.5 we see that a necessary condition for A to have an inverse is that $r(A) = n$. The converse will be shown in 3.5.7.

3.5.3 Uniqueness of the inverse, definition of A^{-1}

Let X and Y be $n \times n$ matrices such that $AX = I$ and $YA = I$. Then $X = Y$ and this matrix is the unique inverse of A.
Proof: We have

$$Y = YI = Y(AX) = (YA)X = IX = X \tag{3.62}$$

Hence $X = Y$ is an inverse of A. But two inverses X and Y of A must satisfy (3.62), and so A can have only one inverse. It is denoted by A^{-1}.

Example 3.5.1 *Show directly that*

$$\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix}$$

Solution:

By direct multiplication,

$$\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}$$

3.5.4 Use of the inverse

3.5.5 Solving $A\underline{x} = \underline{b}$

Consider a system

$$A\underline{x} = \underline{b} \tag{3.63}$$

of n equations for n unknowns. If A has an inverse, then the unique solution is

$$\underline{x} = A^{-1}\underline{b}$$

In the first place, we show that $\underline{x} = A^{-1}\underline{b}$ is a solution to (3.63):

$$A\left(A^{-1}\underline{b}\right) = \left(AA^{-1}\right)\underline{b} = I\underline{b} = \underline{b}$$

Conversely, let $\underline{x} = \underline{u}$ satisfy the equation (3.63), i.e. suppose $A\underline{u} = \underline{b}$. Multiply both sides of this equation on the left by A^{-1}:

$$A^{-1}(A\underline{u}) = \left(A^{-1}A\right)\underline{u} = I\underline{u} = \underline{u} = A^{-1}\underline{b}$$

Thus $\underline{x} = A^{-1}\underline{b}$ is the unique solution to (3.63).

3.5.6 Solving $\underline{y}A = \underline{c}$

If A has an inverse, then for any row vector $\underline{c}$ with n entries, the unique solution to

$$\underline{y}A = \underline{c} \tag{3.64}$$

is

$$\underline{y} = \underline{c}A^{-1}$$

The proof can be derived from 3.5.5 using transposes (see equation 3.9). Alternatively, the proof parallels that of the previous section. The unique solution is found by multiplying both sides of equation (3.64) on the right by A^{-1}.

Example 3.5.2 *Solve*

$$\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix} \underline{x} = \begin{bmatrix} 9 \\ -11 \end{bmatrix}$$

Solution:

Use the result of Example 3.5.1:

$$\underline{x} = \begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ -11 \end{bmatrix} = \begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ -11 \end{bmatrix} = \begin{bmatrix} 122 \\ -51 \end{bmatrix}$$

3.5.7 Necessary and sufficient conditions for A to be invertible

The $n \times n$ matrix A is invertible if, and only if, any of the following equivalent conditions hold.

If (14) holds, necessarily $X = A^{-1}$. If (15) holds, necessarily $Y = A^{-1}$.

1. $r(A) = n$.
2. The columns of A are independent.
3. The rows of A are independent.
4. The homogeneous equation $A\underline{x} = \underline{0}$ has only the trivial solution.
5. The homogeneous equation $\underline{y}A = \underline{0}$ has only the trivial solution.
6. The row REF of A is I.
7. The column REF of A is I (the row REF of A^T is I).
8. For every column vector $\underline{b}$ with n entries $A\underline{x} = \underline{b}$ has a solution.
9. For every column vector $\underline{b}$ with n entries $A\underline{x} = \underline{b}$ has a unique solution.
10. There is a column vector $\underline{b}$ with n entries such that $A\underline{x} = \underline{b}$ has a unique solution.
11. For every row vector $\underline{c}$ with n entries $\underline{y}A = \underline{c}$ has a solution.
12. For every row vector $\underline{c}$ with n entries $\underline{y}A = \underline{c}$ has a unique solution.
13. There is a row vector $\underline{c}$ with n entries such that $\underline{y}A = \underline{c}$ has a unique solution.
14. There is a matrix X such that $AX = I$.
15. There is a matrix Y such that $YA = I$.

Proof

The first seven properties are equivalent from the definition of linear independence, the fact that row-rank is equal to column-rank, and because the row (column) REF of A is I if, and only if, the columns (rows) of A are independent.

Now assume $r(A) = n$. Since the addition of column $\underline{b}$ (row $\underline{c}$) to A does not change the column (row) rank, $A\underline{x} = \underline{b}$ has a solution by Lemma 3.4.21, and similarly so does $\underline{y}A = \underline{c}$. These solutions are necessarily unique on account of the independence of the columns (rows) of A.

Assume next that (8) holds. Then we can solve the n matrix equations $AX_{\bullet j} = I_{\bullet j}$ for the columns $X_{\bullet j}$ and obtain an $n \times n$ matrix X such that $AX = I$. Similarly, if (11) holds we can find an $n \times n$ matrix Y such that $YA = I$.

By No.5 of Exercise 78, if either $AX = I$ for some X or $YA = I$ for some Y, necessarily

$r(A) = n$.

Thus all the above properties (1) - (13) are equivalent.

Finally, suppose (14) holds. Then, as the properties are equivalent, $YA = I$ for some Y. By 3.5.3, $X = Y$ and X is the inverse of A.
A similar argument shows that if (15) holds, necessarily $Y = A^{-1}$.

Example 3.5.3 *Show that none of the matrices*

$$\left[\begin{array}{cc} 2 & -3 \\ 0 & 0 \end{array}\right], \left[\begin{array}{cc} 2 & 0 \\ 1 & 0 \end{array}\right], \left[\begin{array}{cc} 2 & 2 \\ 1 & 1 \end{array}\right]$$

can have inverses.

Solution:

None of these has rank 2.

3.5.8 Reduction method finds A^{-1}, or decides it does not exist

Reduce the augmented array $[A, I]$ to row REF $[A'', I'']$. Then A has an inverse if, and only if, $A'' = I$, in which case $I'' = A^{-1}$. We are using property (14) of 3.5.7.
Note that if at any stage say, $[A', I']$ of the reduction, A' has a zero row, then A^{-1} cannot exist.

Example 3.5.4 *Consider the matrix of example 3.1.7:*

$$A = \left[\begin{array}{ccc} 2 & 5 & 1 \\ -3 & -6 & 0 \\ 4 & 7 & -2 \end{array}\right]$$

We found there that an echelon form has three non-zero rows. Therefore, $r(A) = 3$ and A^{-1} exists. Find A^{-1} by row reduction.

Solution:

Start with

$$\begin{array}{cccccc} 2 & 5 & 1 & 1 & 0 & 0 \\ -3 & -6 & 0 & 0 & 1 & 0 \\ 4 & 7 & -2 & 0 & 0 & 1 \end{array}$$

$$\begin{array}{cccccc} 2 & 5 & 1 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & 1 & 0 \\ 0 & -3 & -4 & -2 & 0 & 1 \end{array} \quad \begin{array}{c} \\ \frac{3}{2}R_1 + R_2 \\ (-2)\,R_1 + R_3 \end{array}$$

$$\begin{array}{cccccc} 2 & 5 & 1 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & 1 & 0 \\ 0 & 0 & -1 & 1 & 2 & 1 \end{array} \quad \begin{array}{c} \\ \\ 2R_2 + R_3 \end{array}$$

$$\begin{array}{cccccc} 2 & 5 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & \frac{2}{3} & 0 \\ 0 & 0 & 1 & -1 & -2 & -1 \end{array} \quad \begin{array}{c} \\ \frac{2}{3}R_2 \\ (-1)\,R_3 \end{array}$$

$$\begin{array}{cccccc} 2 & 5 & 0 & 2 & 2 & 1 \\ 0 & 1 & 0 & 2 & \frac{8}{3} & 1 \\ 0 & 0 & 1 & -1 & -2 & -1 \end{array} \quad \begin{array}{c} -R_3 + R_1 \\ -R_3 + R_2 \\ \\ \end{array}$$

$$\begin{array}{cccccc|l} 2 & 0 & 0 & -8 & -\frac{34}{3} & -4 & -5R_2+R_1 \\ 0 & 1 & 0 & 2 & \frac{8}{3} & 1 & \\ 0 & 0 & 1 & -1 & -2 & -1 & \end{array}$$

$$\begin{array}{cccccc|l} 1 & 0 & 0 & -4 & -\frac{17}{3} & -2 & \frac{1}{2}R_1 \\ 0 & 1 & 0 & 2 & \frac{8}{3} & 1 & \\ 0 & 0 & 1 & -1 & -2 & -1 & \end{array}$$

Hence

$$A^{-1} = \begin{bmatrix} -4 & -\frac{17}{3} & -2 \\ 2 & \frac{8}{3} & 1 \\ -1 & -2 & -1 \end{bmatrix}$$

Exercise 86 *1. Find the inverses of the following matrices (where possible), by reduction to I, and solve the related problems.*

(a)

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$$

(b) i.

$$B = \begin{bmatrix} 3 & 2 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & -1 \end{bmatrix}$$

ii. Find the linear transformation T that maps column $B_{\bullet j}$ of B to the corresponding column $Y_{\bullet j}$ of Y for $j = 1, 2, 3$, where

$$Y = \begin{bmatrix} 1 & 0 & 4 \\ -1 & 1 & -2 \\ 2 & 3 & 1 \end{bmatrix}$$

Hint 87 *You want a 3×3 matrix T such that $Y_{\bullet j} = TB_{\bullet j}$. In fact, $T = YB^{-1}$*

(c)

$$C = \begin{bmatrix} 1 & 6 & 2 \\ -2 & 3 & 5 \\ 7 & 12 & -4 \end{bmatrix}$$

(d) i.

$$D = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 0 & -3 \\ -1 & 4 & 2 \end{bmatrix}$$

ii. Use your inverse to solve

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 3 \\ 2x_1 - 3x_3 &= -1 \\ -x_1 + 4x_2 + 2x_3 &= 2 \end{aligned}$$

iii. Use your inverse to solve

$$\begin{aligned} y_1 + 2y_2 - y_3 &= 4 \\ 3y_1 + 4y_3 &= 5 \\ -y_1 - 3y_2 + 2y_3 &= -1 \end{aligned}$$

2. Prove that an invertible matrix maps linearly independent vectors to linearly independent vectors and state and prove the converse.

3. *Let A and B be $n \times n$ matrices. Show that AB is invertible if, and only if, both A and B are, in which case $(AB)^{-1} = B^{-1}A^{-1}$. Generalize this result to products of three or more matrices.*

4. *Let the $n \times n$ matrix A be invertible. Show that A^T is also invertible and that $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.*

5. *Let A be $n \times n$. As a linear transformation, the mapping $A : \Re^n \to \Re^n$ is one-to-one if, and only if, A maps $\Re^n$ onto the whole of $\Re^n$.*
Prove the statement. (Compare Example 3.4.4).

6. *If M is a matrix that reflects in a plane or line passing through the origin, show algebraically that $M^{-1} = M$.*

   Hint 88 *Use $M^2 = I$ from Exercise 63, No.17*

7. ** Let A be an $m \times n$ matrix of rank k. Show that by applying elementary row operations and elementary column operations to the matrix A we can find an invertible matrices P and Q such that*

$$PAQ = \begin{bmatrix} I_k & O_{k\times(n-k)} \\ O_{(m-k)\times k} & O_{(m-k)\times(n-k)} \end{bmatrix}$$

   Hint 89 *First reduce A to row REF. Now use elementary column operations and the pivot columns to reduce the non-pivot columns to zero columns. Apply suitable permutations to the columns and finally use the results of Exercise 78), No.11.(Compare No.6 of the same exercise).*

8. *In Chapter 2, Exercise 39), No.3 we considered the system $A\underline{x} = \underline{b}$ of two equations for two unknowns (2.32). Show directly that, provided the determinant defined by $|A| = a_{11}a_{22} - a_{21}a_{12}$ is not zero,*

$$B = \frac{1}{|A|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

   satisfies $AB = BA = I$, so that $B = A^{-1}$. Hence derive Cramer's solution of the equations, as found in Chapter 2. (This theory will be further developed in Chapter 4).

9. *Let A be a 3×3 matrix such that $2I - 3A + 5A^2 - 7A^4 = 0$, where 0 is the 3×3 zero matrix. Show that A has an inverse and write it in terms of A.*

   For the solution *see Chapter 4, subsection 4.1.1 and equation (4.1.3).*

   Hint 90 *$A(7A^3 - 5A + 3I) = 2I$*

10. *Find all $n \times n$ invertible A such that $A^2 = A$.*

3.6 Vector spaces: Subspaces of $\Re^m$

Let $\Re^m$ stand for $\Re^{m\times 1}$ (the set of column vectors with m entries) or $\Re^{1\times m}$ (the set of row vectors with m entries). (The context will decide which is meant, or it may not matter, as in Theorem A below, subsection 3.6.4.)
A non-empty subset $\mathcal{S} \subseteq \Re^m$ is said to be a **subspace** of $\Re^m$ (or a **vector space** in $\Re^m$) if whenever $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k$ are in $\Re^m$ it follows that any linear combination of $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k$ is also in $\Re^m$. One says that $\mathcal{S}$ is **closed under linear combinations**.

It is easy to see that $\mathcal{S}$ is a subspace of $\Re^m$ if $\mathcal{S}$ is closed under linear combinations of any two vectors from $\mathcal{S}$. (See Exercise 93, No.1.)

3.6.1 Examples

1. The set $\mathcal{S} = \{\underline{0}\}$ consisting of only the zero vector $\underline{0} \in \Re^{m\times 1}$ is a subspace of $\Re^{m\times 1}$. The whole space $\Re^{m\times 1}$ is a subspace of itself. These are called **trivial** subspaces of $\Re^{m\times 1}$.

2. Let A be an $m \times n$ matrix.

3. The row space of A is a subspace of $\Re^{1\times n}$. We denote this space by $\mathcal{R}(A)$. This is the span of the rows of A.

4. The column space of A is a subspace of $\Re^{m\times 1}$. It is denoted by $\mathcal{C}(A)$. This is the range of the mapping $\underline{x} \to A\underline{x}$.

   The set of solutions $\underline{x}$ of the homogeneous equation $A\underline{x} = \underline{0}$ is a subspace of $\Re^{n\times 1}$. It is called the **null space** of A and is denoted by $\mathcal{N}(A)$.

5. Let A be $n \times n$ and $I = I_n$. A scalar λ such that the matrix $A - lambdaI$ fails to be invertible is called an **eigenvalue** of A (see No.9 in Exercise 101). The null space $\mathcal{N}(A - \lambda I)$ is called the **eigenspace** belonging to λ.
   These are important concepts, both for pure and applied mathematics.

3.6.2 Bases and dimension

We suppose for convenience here that subspaces are contained in $\Re^m = \Re^{m\times 1}$.

3.6.3 Definition of basis

Let $\mathcal{S} \neq \{\underline{0}\}$ be a subspace of $\Re^m$. A set $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ of linearly independent vectors that span $\mathcal{S}$ is called a **basis** of $\mathcal{S}$.

The main results are the following

3.6.4 Theorem A

Let $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ be a basis for the subspace $\mathcal{S} \neq \{\underline{0}\}$ of $\Re^m$. Then

1. The space $\mathcal{S}$ contains at most n linearly independent vectors.

2. If $\underline{b}_2, \ldots, \underline{b}_k$ are vectors in $\mathcal{S}$ that span $\mathcal{S}$ then $k \geq n$.

3. Any two bases of $\mathcal{S}$ contain the same number of vectors.

Proof

1. Suppose that the vectors $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ are linearly independent. By the fundamental theorem in 3.4.22, since they are in the span of the vectors $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$, adding the $\underline{b}_j$ to this set cannot create more than n independent vectors. Thus $k \leq n$. This proves 1.

2. A similar argument shows that if the vectors $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ in $\mathcal{S}$ span the space $\mathcal{S}$, then $k \geq n$, since the independent set $\underline{a}_1, \ldots, \underline{a}_n$ is contained in this span. This proves 2.

3. If the vectors $\underline{b}_1, \ldots, \underline{b}_k$ form a basis of $\mathcal{S}$, then $k \geq n$ and $k \leq n$. Hence $k = n$. This is the statement of 3.

Corollary 91 *A consequence of 1 is yet another proof that a homogeneous system $A\underline{x} = \underline{0}$ of m equations in $n > m$ unknowns has a non-trivial solution. (See 2.2.15).*
For, $\Re^m$ has a basis consisting of the m unit vectors $\underline{e}_j$ and A has $n > m$ columns in $\Re^m$.

3.6.5 Theorem B

A subspace $\mathcal{S} \neq \{\underline{0}\}$ of $\Re^m$ always has a basis.

Proof:

Consider the following process. Let $\underline{a}_1 \neq \{\underline{0}\}$ be a member of $\mathcal{S}$. If $\underline{a}_1$ spans $\mathcal{S}$ the process stops. Otherwise there is some $\underline{a}_2 \notin sp(\underline{a}_1)$. By the fundamental theorem on linear independence, 3.4.22, $\underline{a}_1$ and $\underline{a}_2$ are linearly independent. If $\mathcal{S} = sp(\underline{a}_1, \underline{a}_2)$, stop. Otherwise, continuing in this way, the process must stop at some stage since $\Re^m$ has at most m linearly independent vectors. At which point we have linearly independent vectors $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ which span $\mathcal{S}$. This proves the theorem.

Remark 92 *Let the basis $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ for $\mathcal{S}$ be the columns of the $m \times n$ matrix $A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n]$. What we have found is that $r(A) = n$ and the column space $\mathcal{C}(A) = \mathcal{S}$.*

Definition

The number k of vectors in any basis of $\mathcal{S}$ is called the **dimension** of $\mathcal{S}$, written $k = \dim(\mathcal{S})$. By definition the dimension of the trivial space $\{\underline{0}\}$ is 0.

Exercise 93

1. *Let $\mathcal{S}$ be a non-empty subset of $\Re^m$. Show that it is a subspace if, and only if the following conditions hold: (i). If $\underline{a} \in \mathcal{S}$ and γ is a scalar, then $\gamma\underline{a} \in \mathcal{S}$. (ii). If $\underline{a}$ and $\underline{b}$ are both in $\mathcal{S}$ then $\underline{a} + \underline{b} \in \mathcal{S}$.*

2. *Show that the items listed in 3.6.1 do satisfy the criteria of No.1 to be subspaces.*

3. *Show that the dimension of $\Re^m$ is m.*

4. *Show that the columns of a matrix A are linearly independent if, and only if, $\mathcal{N}(A) = \{\underline{0}\}$. (See 3.56.)*

5. *Let A be an $n \times n$ matrix. Write down three equivalent necessary and sufficient on the dimensions of the row, column and null spaces of A for A to be invertible.*

6. *Let $\mathcal{S}$ and $\mathcal{T}$ be subspaces of $\Re^m$. The sum $\mathcal{S}+\mathcal{T}$ consists of all vectors $\underline{a}+\underline{b}$, where $\underline{a} \in \mathcal{S}$ and $\underline{b} \in \mathcal{T}$.*
   Show that both the intersection $\mathcal{S} \cap \mathcal{T}$ and sum $\mathcal{S} + \mathcal{T}$ are subspaces of $\Re^m$.

7. *Without any reference to elementary row operations use the fundamental theorem 3.4.22 to prove that the column - and row-ranks of an $m \times n$ matrix A are equal.*

   Hint 94 *Use results from Theorem A above and its Corollary 91 and suppose the columns of the $m \times k$ matrix B form a maximum number of independent columns of A. Follow the proof in subsection 3.4.19.*
   Alternatively, you can first assume the m rows of A are independent and show that the column-rank of A is at least m. The general case, column-rank $A \geq$ row-rank A, follows from this.

8. *Show that the column space $\mathcal{C}(A)$ of an $m \times n$ matrix A has a basis consisting of $r(A)$ columns of A. A similar result holds for the row space $\mathcal{R}(A)$.*
   Conclude that the dimension of $\mathcal{C}(A)$ and $\mathcal{R}(A)$ are both $r(A)$.

9. *Let A be an $m \times n$ matrix. Suppose that B is the result of performing an elementary row operation on A (see subsection 2.2.6, Chapter 2). Show that A and B have the same row space. Conclude that if E is an echelon form (or REF) of A, then E and A have the same row space.*

   For example, the rows of the coefficient matrix of example 2.2.5 in Chapter 2 are spanned

by the three rows $\begin{bmatrix} 1 & -1 & 1 & 2 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 3 & 1 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 & -2 \end{bmatrix}$.

Find two rows which span the row space of the matrix A in Exercise 35 of Chapter 2, No.1a.

Hint 95 *Suppose that* $\underline{a}_1, \ldots, \underline{a}_m$ *are the rows of matrix A and let* $y_1, \ldots, y_m$ *and k be scalars. Then,*

$$y_1\underline{a}_1 + \cdots + y_m\underline{a}_m = y_1(\underline{a}_1 + k\underline{a}_2) + (y_2 - y_1k)\underline{a}_2 + \cdots + y_m\underline{a}_m$$

10. *Let* $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$ *span the space* $\mathcal{S}$ *of dimension n. Show that the vectors* $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n$ *are necessarily independent.*

11. *Let independent vectors* $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k$ *be in the space* $\mathcal{S}$ *of dimension n. Show that* $\mathcal{S}$ *has a basis that includes the vectors* $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k$*. (This is implicit in the proof of the above theorem B).*

12. *Let B be a matrix whose columns are in the column space of A, i.e., each* $B_{\bullet j} \in \mathcal{C}(A)$*. Then we know that* $\mathcal{C}(B) \subseteq \mathcal{C}(A)$ *and that* $r(B) \leq r(A)$*. (Why?) Show that* $\mathcal{C}(B) = \mathcal{C}(A)$ *if, and only if,* $r(B) = r(A)$*.*

State and prove similar results for the ***row space*** $\mathcal{R}(A)$ *of A.*

13. ** Show that if matrices A and B have the same size, then*

$$r(A+B) \leq r[A,B] \leq r(A) + r(B)$$

Hint 96 *Use* $\mathcal{C}(A+B) \subseteq \mathcal{C}[A,B] = \mathcal{C}(A) + \mathcal{C}(B)$

14. ** Prove the following exchange theorem, first proved by the 19th Century mathematician H. Grassmann:*

In $\Re^m$ *let* $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ *be linearly independent and suppose that each vector* $\underline{a}_j$ *depends linearly on* $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$*. Then*

(i) $n \leq k$*.*

(ii) At least n vectors among $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$ *can be dropped and replaced with the vectors* $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ *in such a way that the resulting set has the same span as that of* $\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k$*.*

Hint 97 *We may assume* $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n, \underline{b}_1, \underline{b}_2, \ldots, \underline{b}_r$ *is a basis for* $\mathcal{S} = sp(\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_k)$*. Use the fact that the dimension of this space is* $n + r \leq k$*.*

15. ** The following theorem is basic for vector space theory, and is closely related to the result proved in Exercise 52, No. 5:*
Let A be an $m \times n$ *matrix. Show that the null and column spaces of A are related by the equation*

$$dim\mathcal{N}(A) + dim\mathcal{C}(A) = n$$

Hint 98 *Let* $\underline{u}_1, \ldots \underline{u}_k$ *be a basis for the null space* $\mathcal{N}(A)$ *of A. Extend this to a basis* $\underline{u}_1, \ldots \underline{u}_k, \underline{u}_{k+1} \ldots \underline{u}_n$ *for the whole space* $\Re^{n\times 1}$*. Then* $A\underline{u}_{k+1} \ldots A\underline{u}_n$ *is a basis for the column space* $\mathcal{C}(A)$*.*

16. *Deduce the result of Exercise 52, No. 5 from this theorem.*

3.7 Summary of Chapter 3

3.7.1 Matrix notation; columns and rows

(See subsection 3.1.1) An $m \times n$ matrix $A = [a_{ij}]$ is also described by $[A]_{ij} = a_{ij}$.

$\Re^{m \times n}$ denotes the set of $m \times n$ matrices with real number entries.

Row i of the $m \times n$ matrix A is written

$$A_{i\bullet} = \begin{bmatrix} a_{i1} & a_{i2} & .... & a_{in} \end{bmatrix} \quad (i = 1, \cdots, m)$$

Column j is written

$$A_{\bullet j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad (j = 1, \cdots, n)$$

Addition of two matrices of the same size is defined naturally, as is multiplication by a scalar α:

$$[\alpha A]_{ij} = \alpha \, [A]_{ij}$$

If A is $m \times n$ and $\underline{x}$ a column vector with n entries x_i, we have

$$A\underline{x} = \begin{bmatrix} A_{1\bullet}\underline{x} \\ A_{2\bullet}\underline{x} \\ \vdots \\ A_{m\bullet}\underline{x} \end{bmatrix} = A_{\bullet 1}x_1 + A_{\bullet 2}x_2 + ... + A_{\bullet n}x_n$$

Here

$$A_{i\bullet}\underline{x} = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{i1}x_1 & + & \cdots & + & a_{in}x_n \end{bmatrix}$$

and

$$A_{\bullet j}x_j = \begin{bmatrix} a_{1j}x_j \\ a_{2j}x_j \\ \vdots \\ a_{mj}x_j \end{bmatrix}$$

Similarly, if $\underline{y}$ is a row vector with m entries y_j,

$$\underline{y}A = \begin{bmatrix} \underline{y}A_{\bullet 1} & \underline{y}A_{\bullet 2} & \cdots & \underline{y}A_{\bullet n} \end{bmatrix} = y_1A_{1\bullet} + y_2A_{2\bullet} + \cdots + y_mA_{m\bullet}$$

3.7.2 Span, row space, column space

(See (see item 2 in subsection 3.1.1)

The span of the vectors $\underline{a}_1, \ldots, \underline{a}_m$ (all of the same size) is the set of linear combinations of $\underline{a}_1, \ldots, \underline{a}_m$. It is denoted by $sp(\underline{a}_1, \ldots, \underline{a}_m)$.

The row space of a matrix A is the span of its rows (see also exercise 9).

The column space of a matrix A is the span of its columns.

3.7.3 Matrices as mappings

(See section 3.2.
Let A be a fixed $m \times n$ matrix. If $\underline{x}$ is a variable column vector with n entries, then

$$\underline{y} = A\underline{x}$$

expresses $\underline{y}$ as a function of $\underline{x}$; A transforms the vector $\underline{x}$ into the vector $\underline{y}$, so the function is also called a transformation or mapping $\underline{x} \to \underline{y} = A\underline{x}$. The transformation is linear since for all vectors $\underline{x}_1$ and $\underline{x}_2$ and scalars α_1 and α_2,

$$A\left(\underline{x}_1\alpha_1 + \underline{x}_2\alpha_2\right) = \left(A\underline{x}_1\right)\alpha_1 + \left(A\underline{x}_2\right)\alpha_2$$

3.7.4 The product of two matrices

(See subsection 3.3.4)
Let A be an $m \times k$ matrix and B a $k \times n$ matrix. Then the product AB is given by

$$[AB]_{ij} = A_{i\bullet}B_{\bullet j} \quad (i = 1, \ldots, m; j = 1, \ldots, n)$$

Note that AB is only defined when the number of columns of A equals the number of rows of B.

Then AB is $m \times n$ and

$$\begin{aligned} [AB]_{i\bullet} &= A_{i\bullet}B \quad (i = 1, \ldots, m) \\ [AB]_{\bullet j} &= AB_{\bullet j} \quad (j = 1, \ldots, n) \end{aligned}$$

Furthermore,

$$\begin{aligned} A_{i\bullet}B &= a_{i1}B_{1\bullet} + a_{i2}B_{2\bullet} + \cdots + a_{ik}B_{k\bullet} \\ AB_{\bullet j} &= A_{\bullet 1}b_{1j} + A_{\bullet 2}b_{2j} + \cdots + A_{\bullet k}b_{kj} \end{aligned}$$

If $\underline{x}$ is a variable vector with n entries, then the composite transformation $\underline{x} \to A\left(B\underline{x}\right)$ is given by

$$A\left(B\underline{x}\right) = \left(AB\right)\underline{x}$$

More generally, we have the associative law for multiplication: if AB and BC are defined, then so are $\left(AB\right)C$ and $A\left(BC\right)$ and

$$\left(AB\right)C = A\left(BC\right)$$

3.7.5 Linear independence

For this see section 3.4.
The columns of a matrix A are linearly independent if the equation

$$A\underline{x} = \underline{0}$$

has only the trivial solution $\underline{x} = \underline{0}$.

The rows of a matrix A are linearly independent if the equation

$$\underline{y}A = \underline{0}$$

has only the trivial solution $\underline{y} = \underline{0}$.
Elementary row operations preserve linear relations among the columns of A. See 3.4.7.

The column rank of any $m \times n$ matrix A is the maximum number of linearly independent columns. This is the number of pivot columns in a row EF of A. (See subsection 3.4.8).
The row and column ranks of A are equal. (See 3.4.20).
The main theorem on linear independence is 3.4.22): If we add a column $\underline{b}$ with m entries to an $m \times n$ matrix A to form $[A, \underline{b}]$, then A and $[A, \underline{b}]$ have the same column rank if, and only if $A\underline{x} = \underline{b}$ has a solution, i.e., if, and only if, $\underline{b}$ is linearly dependent on the columns of A.

3.7.6 Invertible square matrices

(See subsection 3.5.3.
An $n \times n$ matrix A is invertible if there is a matrix B such that $AB = BA = I$. When this is the case, B is unique and $B = A^{-1}$ is called the inverse of A.

A has an inverse if, and only if, $r(A) = n$ and there are a number of equivalent conditions for this to be the case. See 3.5.7.

3.7.7 Vector spaces

(See section 3.6).
A subspace is a set $\mathcal{S} \subseteq \Re^m$ closed under the taking of linear combinations. One of the main results is the dimension theorem: Any two bases of a space $\mathcal{S}$ have the same number of vectors.

Chapter 4

Determinants and the Cross Product

4.1 Determinants

The determinant of a 1×1 matrix $A = [a]$ is simply $|A| = a$ (not $|a|$).

4.1.1 Determinants of 2×2 matrices

Starting in Chapter 2, Exercise 39 No.3 and continuing in Chapter 3, Exercise 86, No.8 we found that

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

provided that the **determinant** of A,

$$|A| = a_{11}a_{22} - a_{21}a_{12} \tag{4.1}$$

is not zero. We will now see that for 2×2 matrices A,

4.1.2 A has an inverse if, and only if, $|A| \neq 0$.

Proof

Let $d = |A|$ and consider the matrix

$$C = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{4.2}$$

Then, by direct multiplication,

$$CA = AC = \begin{bmatrix} |A| & 0 \\ 0 & |A| \end{bmatrix} = dI \tag{4.3}$$

where $I = I_2$ is the 2×2 identity matrix. Therefore, if $|A| \neq 0$, the matrix

$$B = \frac{1}{|A|}C = \frac{1}{d} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

satisfies

$$BA = AB = I$$

and so B is the inverse of A.
Conversely, suppose that A^{-1}exists. Multiply both sides of equation (4.3) by A^{-1}:

$$A^{-1}(AC) = \left(A^{-1}A\right)C = IC = C = dA^{-1}$$

Now if $d = 0$, then $C = O$, and therefore $A = O$. Obviously, the zero matrix cannot have an inverse, so $d \neq 0$.

4.1.3 Cramer's Rule for two equations in two unknowns

Let us use

$$A^{-1} = \frac{1}{d}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

to solve the system of two equations for two unknowns:

$$A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

We have $A^{-1}(A\underline{x}) = (A^{-1}A)\underline{x} = I\underline{x} = \underline{x} = A^{-1}\underline{b}$, so

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \frac{1}{d}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= \frac{1}{d}\begin{bmatrix} a_{22}b_1 - a_{12}b_2 \\ -a_{21}b_1 + a_{11}b_2 \end{bmatrix} \end{aligned}$$

Or,

$$\begin{aligned} x_1 &= \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} / |A| \\ x_2 &= \begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix} / |A| \end{aligned}$$

Notice how column $\underline{b}$ replaces column $A_{\bullet j}$ in the determinant of A for x_j.
This is known as *Cramer's Rule*, after the 18th Century Swiss mathematician G. Cramer.

4.1.4 Properties of determinants of 2×2 matrices

Although a determinant is a number, equation (4.1) is also called a 2×2 determinant. The following properties hold for 2×2 determinants

1. The determinant of A is the same as the determinant of its transpose:

$$\left|A^T\right| = |A|$$

Proof:If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $A^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$, so
$\left|A^T\right| = a_{11}a_{22} - a_{12}a_{21} = |A|$.

2. Interchanging rows changes the sign of the determinant :

$$\begin{vmatrix} A_{2\bullet} \\ A_{1\bullet} \end{vmatrix} = -\begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \end{vmatrix} = -|A|$$

Proof:

$$\begin{vmatrix} A_{2\bullet} \\ A_{1\bullet} \end{vmatrix} = \begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22} = -|A|$$

3. Multiplying a row by a scalar λ multiplies the determinant by λ:

$$\begin{vmatrix} \lambda A_{1\bullet} \\ A_{2\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ \lambda A_{2\bullet} \end{vmatrix} = \lambda |A|$$

Proof:

$\begin{vmatrix} \lambda A_{1\bullet} \\ A_{2\bullet} \end{vmatrix} = \begin{vmatrix} \lambda a_{11} & \lambda a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \lambda a_{11}a_{22} - a_{21}\lambda a_{12} = \lambda |A|$

Similarly,

$$\begin{vmatrix} A_{1\bullet} \\ \lambda A_{2\bullet} \end{vmatrix} = \lambda |A|$$

4. Let $C = \begin{bmatrix} c_1 & c_2 \end{bmatrix}$, then we have the expansion

$$\begin{vmatrix} A_{1\bullet} + C \\ A_{2\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \end{vmatrix} + \begin{vmatrix} C \\ A_{2\bullet} \end{vmatrix}$$

Proof:

$\begin{vmatrix} A_{1\bullet} + C \\ A_{2\bullet} \end{vmatrix} = \begin{vmatrix} a_{11} + c_1 & a_{12} + c_2 \\ a_{21} & a_{22} \end{vmatrix} = (a_{11} + c_{11})a_{22} - a_{21}(a_{12} + c_{12})$

$= a_{11}a_{22} - a_{21}a_{12} + c_1a_{22} - a_{21}c_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} c_1 & c_2 \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \end{vmatrix} + \begin{vmatrix} C \\ A_{2\bullet} \end{vmatrix}$.

5. If one row is a multiple of another the determinant vanishes:

$$\begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet} \end{vmatrix} = 0$$

Proof: $\begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet} \end{vmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ \lambda a_{11} & \lambda a_{12} \end{bmatrix} = a_{11}\lambda a_{12} - \lambda a_{11}a_{12} = 0.$

6. Adding a multiple of one row to another row leaves the determinant invariant:

$$\begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet} + A_{2\bullet} \end{vmatrix} = |A|$$

Proof:By (4) and (5), $\begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet} + A_{2\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet} \end{vmatrix} + \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \end{vmatrix} = 0 + |A| = |A|.$

7. If A and B are 2×2 matrices, then

$$|AB| = |A|\,|B|$$

Proof:

To better see the argument let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then using properties (2) - (5) we have:

$$\begin{aligned}
|AB| &= \left| \begin{bmatrix} a & b \\ c & d \end{bmatrix} B \right| = \begin{vmatrix} aB_{1\bullet} + bB_{2\bullet} \\ cB_{1\bullet} + dB_{2\bullet} \end{vmatrix} \\
&= \begin{vmatrix} aB_{1\bullet} \\ cB_{1\bullet} + dB_{2\bullet} \end{vmatrix} + \begin{vmatrix} bB_{2\bullet} \\ cB_{1\bullet} + dB_{2\bullet} \end{vmatrix} \\
&= \begin{vmatrix} aB_{1\bullet} \\ cB_{1\bullet} \end{vmatrix} + \begin{vmatrix} aB_{1\bullet} \\ dB_{2\bullet} \end{vmatrix} + \begin{vmatrix} bB_{2\bullet} \\ cB_{1\bullet} \end{vmatrix} + \begin{vmatrix} bB_{2\bullet} \\ dB_{2\bullet} \end{vmatrix} \\
&= 0 + \begin{vmatrix} aB_{1\bullet} \\ dB_{2\bullet} \end{vmatrix} - \begin{vmatrix} cB_{1\bullet} \\ dB_{2\bullet} \end{vmatrix} + 0 \\
&= (ad - bc)\,|B| = |A|\,|B|
\end{aligned}$$

8. The above properties hold for columns in place of rows. In No.1 of Exercise 101 you are asked to formally state and prove this.

4.1.5 3×3 determinants

4.1.6 The search for a formula solving simultaneous equations

Let us try to solve formally

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

The detached array for this system is

$$\begin{array}{cccc} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}$$

Assuming $a_{11} \neq 0$ this is equivalent to

$$\begin{array}{cccc} a_{11} & a_{12} & a_{13} & b_1 \\ 0 & -\frac{a_{21}}{a_{11}}a_{12} + a_{22} & -\frac{a_{21}}{a_{11}}a_{13} + a_{23} & -\frac{a_{21}}{a_{11}}b_1 + b_2 \quad -\frac{a_{21}}{a_{11}}R_1 + R_2 \\ 0 & -\frac{a_{31}}{a_{11}}a_{12} + a_{32} & -\frac{a_{31}}{a_{11}}a_{13} + a_{33} & -\frac{a_{31}}{a_{11}}b_1 + b_3 \quad -\frac{a_{31}}{a_{11}}R_1 + R_3 \end{array}$$

In particular, this means

$$\begin{bmatrix} -\frac{a_{21}}{a_{11}}a_{12} + a_{22} & -\frac{a_{21}}{a_{11}}a_{13} + a_{23} \\ -\frac{a_{31}}{a_{11}}a_{12} + a_{32} & -\frac{a_{31}}{a_{11}}a_{13} + a_{33} \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{a_{21}}{a_{11}}b_1 + b_2 \\ -\frac{a_{31}}{a_{11}}b_1 + b_3 \end{bmatrix}$$

Assuming that

$$\begin{vmatrix} -\frac{a_{21}}{a_{11}}a_{12} + a_{22} & -\frac{a_{21}}{a_{11}}a_{13} + a_{23} \\ -\frac{a_{31}}{a_{11}}a_{12} + a_{32} & -\frac{a_{31}}{a_{11}}a_{13} + a_{33} \end{vmatrix} \neq 0$$

we can use Cramer's rule to find

$$x_2 = \frac{\begin{vmatrix} -\frac{a_{21}}{a_{11}}b_1 + b_2 & -\frac{a_{21}}{a_{11}}a_{13} + a_{23} \\ -\frac{a_{31}}{a_{11}}b_1 + b_3 & -\frac{a_{31}}{a_{11}}a_{13} + a_{33} \end{vmatrix}}{\begin{vmatrix} -\frac{a_{21}}{a_{11}}a_{12} + a_{22} & -\frac{a_{21}}{a_{11}}a_{13} + a_{23} \\ -\frac{a_{31}}{a_{11}}a_{12} + a_{32} & -\frac{a_{31}}{a_{11}}a_{13} + a_{33} \end{vmatrix}} \tag{4.4}$$

Multiply numerator and denominator of (4.4) by $a_{11}^2$and use property 3 of subsection 4.1.4 above:

$$x_2 = \frac{\begin{vmatrix} a_{11}b_2 - a_{21}b_1 & a_{23}a_{11} - a_{21}a_{13} \\ a_{11}b_3 - a_{31}b_1 & a_{33}a_{11} - a_{31}a_{13} \end{vmatrix}}{\begin{vmatrix} a_{11}a_{22} - a_{21}a_{12} & a_{23}a_{11} - a_{21}a_{13} \\ a_{11}a_{32} - a_{31}a_{12} & a_{33}a_{11} - a_{31}a_{13} \end{vmatrix}}$$

Consider the denominator:

$(a_{11}a_{22} - a_{21}a_{12})(a_{33}a_{11} - a_{31}a_{13}) - (a_{11}a_{32} - a_{31}a_{12})(a_{23}a_{11} - a_{21}a_{13})$

$= a_{11}a_{11}a_{22}a_{33} - a_{11}a_{13}a_{22}a_{31} - a_{11}a_{12}a_{21}a_{33} + a_{12}a_{21}a_{31}a_{13}$

$-a_{11}a_{11}a_{23}a_{32} + a_{11}a_{13}a_{21}a_{32} + a_{11}a_{12}a_{23}a_{31} - a_{12}a_{13}a_{21}a_{31}$

$= a_{11}(a_{11}a_{22}a_{33} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31})$.

The numerator for x_2 is found by substituting b_1 for a_{12}, b_2 for a_{22} and b_3 for a_{32}:
$a_{11}(a_{11}b_2a_{33} - a_{13}b_2a_{31} - b_1a_{21}a_{33} - a_{11}a_{23}b_3 + a_{13}a_{21}b_3 + b_1a_{23}a_{31})$.

Finally, we find

$$x_2 = \frac{a_{11}b_2a_{33} - a_{13}b_2a_{31} - b_1a_{21}a_{33} - a_{11}a_{23}b_3 + a_{13}a_{21}b_3 + b_1a_{23}a_{31}}{a_{11}a_{22}a_{33} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31}} \tag{4.5}$$

The denominator is defined as the **determinant** $|A|$ of the 3×3 matrix A:

$$|A| = a_{11}a_{22}a_{33} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} \tag{4.6}$$

$|A|$ is called a 3×3 determinant and it is sometimes written as $\det A$.

Equation (4.5) can be written

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{|A|}$$

Can you guess what $x_1$and x_3 are, if $|A| \neq 0$? This is again Cramer's rule. (See exercise No.7 below).

4.1.7 Discussion on the definition of the 3×3 determinant $|A|$

Each term of (4.6) is of the form

$$\pm a_{1i}a_{2j}a_{3k}$$

The column indices i, j, and k form an arrangement (permutation) ijk of the numbers 1, 2, and 3. In fact all six permutations occur and each has its own **sign** according to the scheme:

$$+123 \quad -321 \quad -213 \quad -132 \quad +312 \quad +231$$

4.1.8 The rule is for finding the sign of ijk

Start with 123 as +. Each time an interchange of two numbers occurs, change the sign as follows:

$+123 \longrightarrow -132$ (interchange 2 and 3)

$-132 \longrightarrow +231$ (interchange 1 and2)

$+231 \longrightarrow -213$ (interchange 1 and 3)

$-213 \longrightarrow +312$ (interchange 2 and 3)

$+312 \longrightarrow -321$ (interchange 1 and 2)

4.1.9 Formal definition of $|A|$

Let $\sigma = (\sigma 1, \sigma 2, \sigma 3)$ denote a typical permutation of $1, 2, 3$ and let $\text{sgn}(\sigma))$ be its sign. Then we can write

$$|A| = \sum_{\sigma} \text{sgn}(\sigma) a_{1\sigma 1} a_{2\sigma 2} a_{3\sigma 3} \tag{4.7}$$

Here the sum extends over all $3! = 6$ permutations σ of 1, 2, 3.

4.1.10 The cofactor $A_{i|j}$ of a_{ij}

The **cofactor** of a_{ij} is defined as

$$A_{i|j} = (-1)^{i+j} \text{ times determinant formed by crossing out row } i \text{ and column } j$$

Example 4.1.1 $A_{1|1} = (-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$,

$A_{1|2} = (-1)^{1+2} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = - \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$,

$A_{2|2} = (-1)^{2+2} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$,

$A_{2|3} = (-1)^{2+3} \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} = - \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$,

$A_{3|1} = (-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$

4.1.11 The sign pattern for the cofactors $A_{i|j}$

Each time we move by one row or by one column we get a change in sign:

$$\begin{array}{ccc} + & - & + \\ - & + & - \\ + & - & + \end{array}$$

4.1.12 Expansion of $|A|$ by a row

Expansion by row 1:

$|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$

$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$

$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

Hence

$$|A| = a_{11}A_{1|1} + a_{12}A_{1|2} + a_{13}A_{1|3}$$

Expansion by row 2:

$|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$

$$-a_{21}(a_{12}a_{33}-a_{13}a_{32})+a_{22}(a_{11}a_{33}-a_{13}a_{31})-a_{23}(a_{11}a_{32}-a_{12}a_{31})$$

$$=-a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}+a_{22}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}-a_{23}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

$$|A|=a_{21}A_{2|1}+a_{22}A_{2|2}+a_{23}A_{2|3}$$

Expansion by row 3:

$|A|=a_{11}a_{22}a_{33}-a_{11}a_{23}a_{32}+a_{12}a_{23}a_{31}-a_{12}a_{21}a_{33}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}$
$=a_{31}\left(a_{12}a_{23}-a_{13}a_{22}\right)-a_{32}\left(a_{11}a_{23}-a_{13}a_{21}\right)+a_{33}\left(a_{11}a_{22}-a_{12}a_{21}\right)$

$$=a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}-a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}+a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

Hence:

$$|A|=a_{31}A_{3|1}+a_{32}A_{3|2}+a_{33}A_{3|3}$$

4.1.13 Expansion by row i

$$|A|=a_{i1}A_{i|1}+a_{i2}A_{i|2}+a_{i3}A_{i|3} \tag{4.8}$$

Example 4.1.2 *Let*

$$A=\begin{bmatrix} 1 & 3 & -1 \\ 2 & 4 & 0 \\ -2 & 5 & -3 \end{bmatrix}$$

Find $|A|$ by expanding by the second row.

Solution:

$$\begin{aligned} |A| &= a_{21}A_{2|1}+a_{22}A_{2|2}+a_{23}A_{2|3} \\ &= (2)(-1)^{2+1}\begin{vmatrix} 3 & -1 \\ 5 & -3 \end{vmatrix}+(4)(-1)^{2+2}\begin{vmatrix} 1 & -1 \\ -2 & -3 \end{vmatrix} \\ &= (-2)(-4)+(4)(-5)=-12 \end{aligned}$$

4.1.14 Properties of 3×3 determinants

1. The determinant of A and its transpose are equal:

$$|A|=\left|A^T\right|$$

Proof:

$|A|=a_{11}a_{22}a_{33}-a_{13}a_{22}a_{31}-a_{12}a_{21}a_{33}-a_{11}a_{23}a_{32}+a_{13}a_{21}a_{32}+a_{12}a_{23}a_{31}$

$=a_{11}a_{22}a_{33}-a_{31}a_{22}a_{13}-a_{21}a_{12}a_{33}-a_{11}a_{32}a_{23}+a_{21}a_{32}a_{13}+a_{31}a_{12}a_{23}.$

If $B=A^T$ then $b_{ij}=a_{ji}$ by definition, so that $|A|$ is equal to

$b_{11}b_{22}b_{33}-b_{13}b_{22}b_{31}-b_{12}b_{21}b_{33}-b_{11}b_{23}b_{32}+b_{12}b_{23}b_{31}+b_{13}b_{21}b_{32}$

$= |B|$

This proves the result.

Note 99 *The idea of the proof can be seen as follows:*
Consider some term of the full expansion of $|A|$, say $-a_{12}a_{21}a_{33}$, or $+a_{13}a_{21}a_{32}$. In the first case, $-a_{12}a_{21}a_{33} = -a_{21}a_{12}a_{33} = -b_{12}b_{21}b_{33}$. In the second case, $+a_{13}a_{21}a_{32} = +a_{21}a_{32}a_{13} = +b_{12}b_{23}b_{31}$. In both cases we get a typical term of $|B|$ with its correct sign.

2. Interchanging two rows changes the sign of the determinant.

   Proof:

   Consider, for example, interchanging rows 2 and 3. Expand $|A|$ by row 1:

$$\begin{aligned} |A| &= a_{11}A_{1|1} + a_{12}A_{1|2} + a_{13}A_{1|3} \\ &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \end{aligned}$$

   Hence, if we interchange $A_{2\bullet}$ and $A_{3\bullet}$, the 2×2 determinants will change sign and so will $|A|$.

3. Multiplying a row of A by a scalar multiplies the determinant by the same scalar:

$$\begin{vmatrix} \lambda A_{1\bullet} \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ \lambda A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \\ \lambda A_{3\bullet} \end{vmatrix} = \lambda |A|$$

   Proof:

   Consider for example multiplying row 2 by λ. Expand the result by row 2:

$$\begin{vmatrix} A_{1\bullet} \\ \lambda A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} = (\lambda a_{21})A_{2|1} + (\lambda a_{22})A_{2|2} + (\lambda a_{23})A_{2|3} = \lambda \left(a_{21}A_{2|1} + a_{22}A_{2|2} + a_{23}A_{2|3}\right) = \lambda |A|$$

   Less elegantly, this can also be seen by expanding by row 1:

$$\begin{aligned} \begin{vmatrix} A_{1\bullet} \\ \lambda A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} &= a_{11}\begin{vmatrix} \lambda a_{22} & \lambda a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} \lambda a_{21} & \lambda a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} \lambda a_{21} & \lambda a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= \lambda |A| \end{aligned}$$

   We have used the fact that each cofactor gets multiplied by λ.

4. Let $B = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}$, then

$$\begin{vmatrix} A_{1\bullet} + B \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} + \begin{vmatrix} B \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} = |A| + \begin{vmatrix} B \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix}$$

   Proof:

$$\begin{aligned} \begin{vmatrix} A_{1\bullet} + B \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} &= (a_{11} + b_1)A_{1|1} + (a_{12} + b_2)A_{1|2} + (a_{13} + b_3)A_{1|3} \\ &= a_{11}A_{1|1} + a_{12}A_{1|2} + a_{13}A_{1|3} + b_1A_{1|1} + b_2A_{1|2} + b_3A_{1|3} \\ &= \begin{vmatrix} A_1 \\ A_2 \\ A_3 \end{vmatrix} + \begin{vmatrix} B \\ A_2 \\ A_3 \end{vmatrix} \end{aligned}$$

Similar results apply if the row B is added to another row, e.g.

$$\begin{vmatrix} A_{1\bullet} \\ A_{2\bullet}+B \\ A_{3\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} + \begin{vmatrix} A_{2\bullet} \\ B \\ A_{3\bullet} \end{vmatrix}$$

Using property (2) this can be proved directly from the first result as follows

$$\begin{vmatrix} A_{1\bullet} \\ A_{2\bullet}+B \\ A_{3\bullet} \end{vmatrix} = -\begin{vmatrix} A_{2\bullet}+B \\ A_{1\bullet} \\ A_{3\bullet} \end{vmatrix} = -\begin{vmatrix} A_{2\bullet} \\ A_{1\bullet} \\ A_{3\bullet} \end{vmatrix} - \begin{vmatrix} B \\ A_{1\bullet} \\ A_{3\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} + \begin{vmatrix} A_{2\bullet} \\ B \\ A_{3\bullet} \end{vmatrix}$$

5. If one row of A is a multiple of another then, $|A| = 0$.
Proof

First suppose that two rows are equal, say $A_{1\bullet} = A_{2\bullet}$. Then interchanging rows 1 and 2 of A changes the sign of the determinant by property (2). Hence $|A| = -|A|$ and so $|A| = 0$. Now let $A_{1\bullet} = \lambda A_{2\bullet}$. By property (3), $|A|$ is λ times a determinant with equal rows 1 and 2 and so vanishes.

6. Adding a multiple of one row to a different row leaves $|A|$ unchanged, e.g.

$$\begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet}+A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} = |A|$$

Proof:

Use property 5:

$$\begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet}+A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} + \begin{vmatrix} A_{1\bullet} \\ \lambda A_{1\bullet} \\ A_{3\bullet} \end{vmatrix} = \begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix} + 0 = |A|$$

7. If A and B are 3×3 matrices,

$$* \, |AB| = |A| \, |B| \tag{4.9}$$

* We will not prove this but note that the proof follows that for 2×2 determinants.

8. Any of the above properties (2) - (6) are true for columns in place of rows.

Proof:

This follows from property (1), namely that $|A| = \left|A^T\right|$.
In particular, we have:

(a) The expansion of $|A|$ by column j:

$$|A| = A_{1|j}a_{1j} + A_{2|j}a_{2j} + a_{3j}A_{3|j}a_{2j}$$

(b) Interchanging two columns changes the sign of the determinant.

(c) Multiplying a column of A by a scalar multiplies the determinant by the same scalar.

(d) If one column of A is a multiple of another, then $|A| = 0$.

(e) Adding a multiple of one column to another leaves $|A|$ unchanged.

Example 4.1.3 *Let*

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 4 & 0 \\ -2 & 5 & -3 \end{bmatrix}$$

Find $|A|$ by expanding by the third column.

Solution:

$$\begin{aligned} |A| &= A_{1|3}a_{13} + A_{2|3}a_{23} + A_{3|3}a_{33} \\ &= (-1)^{1+3}\begin{vmatrix} 2 & 4 \\ -2 & 5 \end{vmatrix}(-1) + A_{2|3}(0) + (-1)^{3+3}\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix}(-3) \\ &= 18(-1) + (-2)(-3) = -12 \end{aligned}$$

Simplification:

We may also simplify the evaluation of the determinant by adding a multiple of a row to another:

$$\begin{vmatrix} 1 & 3 & -1 \\ 2 & 4 & 0 \\ -5 & -4 & 0 \end{vmatrix} \qquad (-3)R_1 + R_3$$

Expanding by the third column gives

$$|A| = (-1)^{1+3}\begin{vmatrix} 2 & 4 \\ -5 & -4 \end{vmatrix}(-1) = -12$$

4.1.15 Fundamental properties of cofactors

For rows:

$$\begin{aligned} a_{i1}A_{j|1} + a_{i2}A_{j|2} + a_{i3}A_{j|3} &= |A| \text{ if } i = j \\ a_{i1}A_{j|1} + a_{i2}A_{j|2} + a_{i3}A_{j|3} &= 0 \text{ if } i \neq j \end{aligned} \tag{4.10}$$

For columns:

$$\begin{aligned} A_{1|i}a_{1j} + A_{2|i}a_{2j} + A_{3|i}a_{2j} &= |A| \text{ if } i = j \\ A_{1|i}a_{1j} + A_{2|i}a_{2j} + A_{3|i}a_{2j} &= 0 \text{ if } i \neq j \end{aligned} \tag{4.11}$$

Proof

We know that if $i = j$ in (4.10) the left-hand side is just the expansion of $|A|$ by row i. Now let $i \neq j$, say $i = 2$ and $j = 1$. Then

$$\begin{aligned} & a_{21}A_{1|1} + a_{22}A_{1|2} + a_{23}A_{1|3} \\ = \; & a_{21}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{22}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{23}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \end{aligned}$$

This is the expansion of

$$\begin{vmatrix} A_{2\bullet} \\ A_{2\bullet} \\ A_{3\bullet} \end{vmatrix}$$

by its first row and so vanishes.
The result for columns can be derived from that for rows.

4.1.16 The adjoint adj(A) of A and the inverse A^{-1}

The **adjoint** of the $n \times n$ matrix A is the $n \times n$ matrix adj(A), where

$$[\text{adj}\,(A)]_{ij} = A_{j|i}$$

This means that we can first find the matrix B with $[B]_{ij} = A_{i|j}$ and then adj$(A) = B^T$.

4.1.17 Properties (4.10) and (4.11) as matrix products

Properties (4.10) and (4.11) just say

$$A\,\text{adj}\,(A) = |A|\,I = \text{adj}\,(A)\,A \tag{4.12}$$

where $I = I_3$ is the 3×3 identity matrix.

Remark 100 *For a* 2×2 *matrix* $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, *the adjoint is adj*$(A) = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$.

This was found in equation (4.2) and the property (4.12) was found in equation (4.3).

4.1.18 Corollary: condition for A to be invertible. Formula for the inverse

The 3×3 matrix A is invertible if, and only if, $|A| \neq 0$, in which case

$$A^{-1} = \frac{1}{|A|}\text{adj}\,(A) \tag{4.13}$$

(For the definition of 'invertible', see Chapter 3, subsection 3.5.2).

Proof

If $|A| \neq 0$, then equation (4.12) shows that

$$A\left(\frac{1}{|A|}\text{adj}\,(A)\right) = I = \left(\frac{1}{|A|}\text{adj}\,(A)\right)A$$

In other words, (4.13) holds.

Conversely, suppose that A^{-1} exists. Let an elementary row operation be done on A to obtain A'.
By properties (2), (3) and (6) we see that $|A| \neq 0$ if, and only if, $|A'| \neq 0$. As A^{-1} exists, its row REF is I. Since $|I| = 1$, it follows that $|A| \neq 0$.

Example 4.1.4 *Let* $A = \begin{bmatrix} 2 & 4 & 3 \\ 0 & 1 & -1 \\ 3 & 5 & 7 \end{bmatrix}$

If A^{-1} *exists, find it.*

Solution:

$|A| = a_{21}A_{2|1} + a_{22}A_{2|2} + a_{23}A_{2|3}$
$= 0 + \begin{vmatrix} 23 \\ 37 \end{vmatrix} - (-1)\begin{vmatrix} 24 \\ 35 \end{vmatrix} = 3$
As $|A| \neq 0$, A^{-1}exists. We find

$$A_{1|1} = \begin{vmatrix} 1 & -1 \\ 5 & 7 \end{vmatrix} = 12,\ A_{1|2} = -\begin{vmatrix} 0 & -1 \\ 3 & 7 \end{vmatrix} = -3,\ A_{1|3} = \begin{vmatrix} 0 & 1 \\ 3 & 5 \end{vmatrix} = -3$$

$$A_{2|1} = -\begin{vmatrix} 4 & 3 \\ 5 & 7 \end{vmatrix} = -13,\ A_{2|2} = \begin{vmatrix} 2 & 3 \\ 3 & 7 \end{vmatrix} = 5,\ A_{2|3} = -\begin{vmatrix} 2 & 4 \\ 3 & 5 \end{vmatrix} = 2$$

$$A_{3|1} = \begin{vmatrix} 4 & 3 \\ 1 & -1 \end{vmatrix} = -7,\ A_{3|2} = -(-2) = 2,\ A_{3|3} = \begin{vmatrix} 2 & 4 \\ 0 & 1 \end{vmatrix} = 2$$

Hence

$$\begin{aligned} A^{-1} &= \frac{1}{|A|}\text{adj}A = \frac{1}{|A|}\begin{bmatrix} A_{1|1} & A_{1|2} & A_{1|3} \\ A_{2|1} & A_{2|2} & A_{2|3} \\ A_{3|1} & A_{3|2} & A_{3|3} \end{bmatrix}^T \\ &= \frac{1}{3}\begin{bmatrix} 12 & -3 & -3 \\ -13 & 5 & 2 \\ -7 & 2 & 2 \end{bmatrix}^T \\ &= \frac{1}{3}\begin{bmatrix} 12 & -13 & -7 \\ -3 & 5 & 2 \\ -3 & 2 & 2 \end{bmatrix} \end{aligned}$$

4.2 Higher Order Determinants

Let A be an $n \times n$ matrix. As with three numbers (see subsection 4.1.7), a permutation $(\sigma 1, \sigma 2, ..., \sigma n)$ of $1, 2, \cdots, n$ is called **even** if it can be obtained from

$1, 2, \cdots, n$ by an even number of interchanges, otherwise odd.

sign $(\sigma) = +1$ if σ is even, and $\text{sgn}(\sigma) = -1$ if σ is odd.

Following the formal definition for 3×3 determinants 4.7, for a general $n \times n$ matrix A, its determinant $|A|$ is given by

$$|A| = \sum_{\sigma} \text{sgn}(\sigma) a_{1\sigma 1} a_{2\sigma 2} ... a_{n\sigma n}$$

where the sum extends over all $n!$ permutations σ of the numbers $1, 2, \cdots, n$.

Recall that $n! = (1)(2)(3)\cdots(n-1)(n)$ and is called n factorial.

As with 2×2 and 3×3 determinants (4.1.10 above), if A is an $n \times n$ matrix, the cofactor $A_{i|j}$ of a_{ij} is given by

$A_{i|j} = (-1)^{i+j}$ times the determinant formed by crossing out row i and column j

For a 4×4 determinant the sign pattern for the cofactors is

$$\begin{matrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{matrix}$$

4.2.1 All the properties of 2×2 and 3×3 determinants hold for $n \times n$ determinants

So properties in 4.1.14, 4.1.13 (equation 4.8) carry over to $n \times n$ determinants. In particular, the formula $|AB| = |A|\,|B|$ and equation (4.12) hold as well as the statement on the inverse in Corollary 4.1.18 and equation (4.13).

Example 4.2.1 *Find* $|A|$ *if*

$$A = \begin{bmatrix} 3 & -2 & -1 & 2 \\ 1 & -2 & 4 & 0 \\ -1 & 3 & 1 & -3 \\ 0 & 1 & -1 & 2 \end{bmatrix}$$

Solution:

$$|A| = \begin{vmatrix} 3 & -2 & -3 & 6 \\ 1 & -2 & 2 & 4 \\ -1 & 3 & 4 & -9 \\ 0 & 1 & 0 & 0 \end{vmatrix}$$

Here we have added column 2 to column 3 and (-2)(Column 2) to column 4. Expanding by row 4:

$$\begin{aligned} |A| &= (-1)^{4+2} \begin{vmatrix} 3 & -3 & 6 \\ 1 & 2 & 4 \\ -1 & 4 & -9 \end{vmatrix} \\ &= \begin{vmatrix} 0 & -3 & 0 \\ 3 & 2 & 8 \\ 3 & 4 & -1 \end{vmatrix} = (-1)^{1+2}(-3) \begin{vmatrix} 3 & 8 \\ 3 & -1 \end{vmatrix} = -81 \end{aligned}$$

We have added column 2 to column 1 and 2(column 2) to column 3 then expanded by row 1.

Exercise 101 ***Partial Solutions***

1. *Using property (1) for* 2×2 *determinants, state and prove similar results for columns in (2)-(6).*

2. *Evaluate the determinants of the matrices in Chapter 3, Exercise 86, No. 1. Find the inverses of the matrices by the adjoint method if they exist.*

   The values of the the determinants are:

   $$|A| = \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} = 1$$

   $$|B| = \begin{vmatrix} 3 & 2 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & -1 \end{vmatrix} = -6$$

   $$|Y| = \begin{vmatrix} 1 & 0 & 4 \\ -1 & 1 & -2 \\ 2 & 3 & 1 \end{vmatrix} = -13$$

   $$|T| = \begin{vmatrix} \frac{1}{3} & -\frac{1}{3} & -\frac{13}{3} \\ -\frac{1}{3} & \frac{5}{6} & \frac{10}{3} \\ \frac{2}{3} & \frac{5}{6} & \frac{4}{3} \end{vmatrix} = \frac{13}{6}$$

   $$|C| = \begin{vmatrix} 1 & 6 & 2 \\ -2 & 3 & 5 \\ 7 & 12 & -4 \end{vmatrix} = 0$$

   $$|D| = \begin{vmatrix} 1 & 3 & -1 \\ 2 & 0 & -3 \\ -1 & 4 & 2 \end{vmatrix} = 1$$

3. *Evaluate the determinants*

(a) $\begin{vmatrix} 1 & -3 & 0 & -2 \\ 3 & -12 & -2 & -6 \\ -2 & 10 & 2 & 5 \\ -1 & 6 & 1 & 3 \end{vmatrix}$

(b) $\begin{vmatrix} -3 & 7 & 8 & 4 \\ 2 & 1 & 0 & 3 \\ 6 & -1 & -2 & -1 \\ -3 & 4 & 2 & 11 \end{vmatrix}$

(c) $\begin{vmatrix} -3 & 7 & 8 & 4 & -7 \\ 2 & 1 & 0 & 3 & -1 \\ 6 & -1 & -2 & -1 & 1 \\ -3 & 4 & 2 & 11 & -4 \\ -3175 & -\frac{25}{2} & 17 & 59 & 12 \end{vmatrix}$

4. *Verify that the adjoint of a 2×2 matrix is given by equation (4.2).*

5. *Let A be a 3×3 matrix and k a scalar. Show that $|kA| = k^3 |A|$. What do you think the general result is?*

6. *Suppose the $n \times n$ matrix A is not invertible. Show that $A(adj\ A)$ is the zero matrix.*

7. *State and prove Cramer's rule for 3 equations in 3 unknowns. Do the same for n equations in n unknowns using the fact that higher order determinants have the same properties as 2 and 3 determinants.*

8. (a) *Let $A = [a_{ij}]$ be a 3×3 matrix and β a scalar. Replace the entry a_{23} with β and let the resulting matrix be B. Show that $|B| = |A| + (\beta - a_{23})\, A_{2|3}$, and find a criterion for B to be invertible.*

   (b) *Generalize the previous result.*

9. *Let A be a 3×3 matrix. A scalar λ such that $A - \lambda I$ is not invertible is called an **eigenvalue** of A (see No.5 in Examples 3.6.1). This means that $|A - \lambda I| = 0$. Show that λ must satisfy a polynomial equation $f(\lambda) = 0$ of degree 3 (called the characteristic polynomial of A). Find this polynomial for the matrix*

$$A = \begin{bmatrix} 10 & -12 & -1 \\ 5 & -6 & -1 \\ -11 & 12 & 0 \end{bmatrix}$$

   as well as the eigenvalues and three corresponding eigenvectors.

   ***Solution**:*
   $f(\lambda) = \lambda^3 - 4\lambda^2 + \lambda + 6$. The eigenvalues are $\lambda = 2$, $\lambda = 3$ and $\lambda = -1$. Eigenvectors belonging to these eigenvalues are respectively
   $\begin{bmatrix} \frac{4}{3} & 1 & -\frac{4}{3} \end{bmatrix}^T$, $\begin{bmatrix} \frac{3}{2} & 1 & -\frac{3}{2} \end{bmatrix}^T$, *and* $\begin{bmatrix} \frac{7}{6} & 1 & \frac{5}{6} \end{bmatrix}^T$.

10. *Let $|A|$ be a 3×3 determinant. We can permute the rows $A_{1\bullet}$, $A_{2\bullet}$ and $A_{3\bullet}$ in six ways. For which of these does the determinant keep its sign and for which does it change its sign?*
    State and prove a similar result for the columns $A_{\bullet 1}$, $A_{\bullet 2}$ and $A_{\bullet 3}$ of A.

4.3 The Cross Product $\underline{a} \times \underline{b}$

In this section all vectors have three entries and, as in the notation of Chapter 1, will normally be row vectors. The unit vectors $\underline{i} = (1, 0, 0)$, $\underline{j} = (0, 1, 0)$ and $\underline{k} = (0, 0, 1)$ were introduced in

subsection 1.1.17.

Let $\underline{a} = (a_1, a_2, a_3)$ and $\underline{b} = (b_1, b_2, b_3)$ are two vectors. Their **cross product** is defined as

$$\begin{aligned} \underline{a} \times \underline{b} &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \underline{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \underline{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \underline{k} \\ &= \left(\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}, \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \right) \end{aligned} \tag{4.14}$$

This is sometimes called the **vector product** and is used constantly in mechanics. Treating $\underline{i}$, $\underline{j}$ and $\underline{k}$ as though they are scalars (which they are not), the cross product can be remembered by the formula

$$\underline{a} \times \underline{b} = \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Expanding this 'determinant' by its first row gives (4.14).
If $\underline{a}$ and $\underline{b}$ are written as columns the convention is to write

$$\underline{a} \times \underline{b} = \begin{vmatrix} \underline{i} & a_1 & b_1 \\ \underline{j} & a_2 & b_2 \\ \underline{k} & a_3 & b_3 \end{vmatrix}$$

but we get the same result as before.

Example 4.3.1

$$\begin{aligned} (-1,2,4) \times (2,-3,1) &= \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ -1 & 2 & 4 \\ 2 & -3 & 1 \end{vmatrix} \\ &= \begin{vmatrix} 2 & 4 \\ -3 & 1 \end{vmatrix} \underline{i} - \begin{vmatrix} -1 & 4 \\ 2 & 1 \end{vmatrix} \underline{j} + \begin{vmatrix} -1 & 2 \\ 2 & -3 \end{vmatrix} \underline{k} \\ &= (14, 9, -1) \end{aligned}$$

Example 4.3.2 *A characteristic property of $\underline{i}$, $\underline{j}$ and $\underline{k}$*

$$\underline{i} \times \underline{j} = \underline{k} \qquad \underline{j} \times \underline{k} = \underline{i} \qquad \underline{k} \times \underline{i} = \underline{j}$$

To see these results, consider

$$\begin{aligned} \underline{i} \times \underline{j} &= \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} \\ &= \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} \underline{i} - \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} \underline{j} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \underline{k} = \underline{k} \end{aligned}$$

$$\begin{aligned} \underline{j} \times \underline{k} &= \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \\ &= \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \underline{i} - \begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix} \underline{j} + \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} \underline{k} = \underline{i} \end{aligned}$$

$$\begin{aligned} \underline{k} \times \underline{i} &= \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix} \\ &= \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} \underline{i} - \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \underline{j} + \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} \underline{k} = \underline{j} \end{aligned}$$

4.3.1 Properties of the cross product

The cross product (4.14) is quite different from the dot product $\underline{a}\cdot\underline{b}$ in that it is a **vector** while the dot product is a **scalar**. Nevertheless, some properties are shared with the dot product.

For all vectors $\underline{a}$, $\underline{b}$ and $\underline{c}$ and scalars β and γ:

1.
$$\underline{b}\times\underline{a} = -(\underline{a}\times\underline{b})$$

Proof:

In $\underline{b}\times\underline{a}$ we interchange the rows in the cofactors defining $\underline{a}\times\underline{b}$. Since these all change sign, the result follows.

2. (The triple scalar product)
$$\underline{c}\cdot(\underline{a}\times\underline{b}) = \underline{b}\cdot(\underline{c}\times\underline{a}) = \underline{a}\cdot(\underline{b}\times\underline{c}) \tag{4.15}$$

Proof:

First note that

$$\begin{aligned}\underline{c}\cdot(\underline{a}\times\underline{b}) &= c_1\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} c_2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} c_3 \\ &= \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}\end{aligned} \tag{4.16}$$

Here we have expanded the determinant by its first row. The other properties follow by rearranging the rows of the above determinant according to the permutations *bca* and *abc* of *cab*. These are all even and so the other determinants are all equal. (See Exercise 101, No.10).

3.
$$\underline{a}\cdot(\underline{a}\times\underline{b}) = \underline{b}\cdot(\underline{a}\times\underline{b}) = 0$$

Proof:

The statement $\underline{a}\cdot(\underline{a}\times\underline{b}) = 0$ follows by putting $\underline{c} = \underline{a}$ in the determinant of (4.16). The resulting determinant is zero as two of its rows are the same. The result $\underline{b}\cdot(\underline{a}\times\underline{b}) = 0$ is similar. Geometrically, this means that if $\underline{a}\times\underline{b}$ is non-zero it is perpendicular to both $\underline{a}$ and $\underline{b}$.

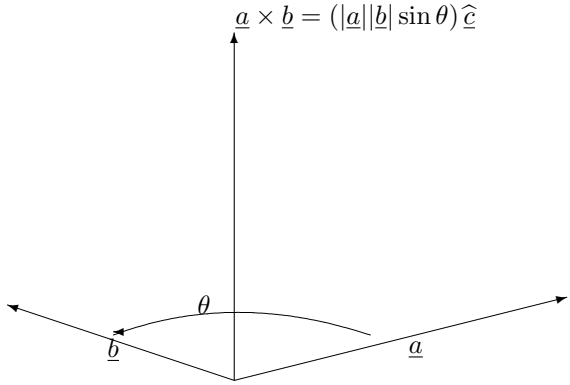
4.
$$\beta(\underline{a}\times\underline{b}) = (\beta\underline{a})\times\underline{b} = \underline{a}\times(\beta\underline{b})$$

5.
$$(\beta+\gamma)(\underline{a}\times\underline{b}) = \beta(\underline{a}\times\underline{b}) + \gamma(\underline{a}\times\underline{b})$$

6.
$$(\alpha\underline{a}+\beta\underline{b})\times\underline{c} = (\alpha\underline{a}\times\underline{c}) + (\beta\underline{b}\times\underline{c})$$

7. If either $\underline{a} = \underline{0}$ or $\underline{b} = \underline{0}$ then $\underline{a}\times\underline{b} = \underline{0}$. Otherwise, $\underline{a}\times\underline{b} = \underline{0}$ if, and only if, $\underline{a}$ and $\underline{b}$ are parallel. (See Exercise 102 No. 7)

Figure 4.1



4.3.2 Geometric meaning of the cross product $\underline{a} \times \underline{b}$

Suppose that $0 < \theta < \pi$ is the angle between the non-zero and non-parallel vectors $\underline{a}$ and $\underline{b}$. Let $\widehat{\underline{c}}$ be the unit vector perpendicular to $\underline{a}$ and $\underline{b}$ with direction found by using the corkscrew rule: rotate from $\underline{a}$ to $\underline{b}$ through θ, then the direction of $\widehat{\underline{c}}$ goes the way a corkscrew would go. See Chapter 1, subsection 1.1.1 and Figure 4.1.

Geometric Interpretation

$$\underline{a} \times \underline{b} = (|\underline{a}||\underline{b}| \sin\theta)\,\widehat{\underline{c}} \tag{4.17}$$

Proof of equation (4.17):

We first show that

$$|\underline{a} \times \underline{b}|^2 + (\underline{a} \cdot \underline{b})^2 = |\underline{a}|^2 |\underline{b}|^2$$

We have

$$\begin{aligned}
& |\underline{a} \times \underline{b}|^2 + (\underline{a} \cdot \underline{b})^2 \\
= & \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}^2 + (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 \\
= & (a_2 b_3 - a_3 b_2)^2 + (a_1 b_3 - a_3 b_1)^2 + (a_1 b_2 - a_2 b_1)^2 + (a_1 b_1 + a_2 b_2 + a_3 b_3)^2 \\
= & a_2^2 b_3^2 + a_3^2 b_2^2 + a_1^2 b_3^2 + a_3^2 b_1^2 + a_1^2 b_2^2 + a_2^2 b_1^2 + a_1^2 b_1^2 + a_2^2 b_2^2 + a_3^2 b_3^2 \\
= & \left(a_3^2 + a_2^2 + a_1^2\right)\left(b_3^2 + b_1^2 + b_2^2\right) = |\underline{a}|^2 |\underline{b}|^2
\end{aligned}$$

It follows that

$$\frac{|\underline{a} \times \underline{b}|^2}{|\underline{a}|^2 |\underline{b}|^2} = 1 - \frac{(\underline{a} \cdot \underline{b})^2}{|\underline{a}|^2 |\underline{b}|^2} = 1 - \cos^2\theta = \sin^2\theta$$

Hence $|\underline{a} \times \underline{b}| = |\underline{a}||\underline{b}| \sin\theta$. Since $\underline{a} \times \underline{b}$ is parallel to $\widehat{\underline{c}}$, we have $\underline{a} \times \underline{b} = \pm\,(|\underline{a}||\underline{b}| \sin\theta)\,\widehat{\underline{c}}$.
If $\underline{a} = \underline{i}$ and $\underline{b} = \underline{j}$, then from Example 4.3.2 we have from the Corkscrew Rule $\widehat{\underline{c}} = \underline{k}$ and $\underline{i} \times \underline{j} = +\underline{k}$. This suggests that the corkscrew rule applied to any cross product $\underline{a} \times \underline{b}$ gives the same direction as $\widehat{\underline{c}}$, and we will be satisfied with this. Equation (4.17) now follows.

From Example 4.3.2 we see how the geometric interpretation gives, besides $\underline{i} \times \underline{j} = \underline{k}$, also the other products $\underline{j} \times \underline{k} = \underline{i}$ and $\underline{k} \times \underline{i} = \underline{j}$.
Compare this with 1.1.1 on right-handed systems in Chapter 1.

Exercise 102 *1. Find* $(-1, 5, 7) \times (2, -3, 4)$ *and* $(-2, 4, -3) \times (6, -1, 2)$.

Solution:

$(-1, 5, 7) \times (2, -3, 4) = (41, 18, -7)$
$(-2, 4, -3) \times (6, -1, 2) = (5, -14, -22)$

2. *Find formulae for $\underline{i} \times \underline{a}$, $\underline{j} \times \underline{a}$ and $\underline{k} \times \underline{a}$.*

3. *Consider three points A, B and C in space. If they are not collinear they form a triangle. Show that its area is $\frac{1}{2}|\underline{AB} \times \underline{AC}|$.*

4. *Find the area of the triangle with vertices $A = (3, 4, 5)$, $B = (4, 4, 6)$, $C = (5, 5, 4)$.*

5. *Show that if the non-zero vectors $\underline{a}$ and $\underline{b}$ are at right angles then $|\underline{a} \times \underline{b}| = |\underline{a}||\underline{b}|$. What of the converse?*

6. *Let $\underline{l}$ and $\underline{m}$ be unit vectors and put $\underline{n} = \underline{l} \times \underline{m}$. Using the geometric interpretation of the cross product show that $\underline{l}$, $\underline{m}$ and $\underline{n}$ form a right-handed system (see 1.1.1). In fact,*

$$\underline{l} \times \underline{m} = \underline{n} \qquad \underline{m} \times \underline{n} = \underline{l} \qquad \underline{n} \times \underline{l} = \underline{m}$$

Compare this with Example 4.3.2.

7. *Complete the proofs of the properties 4.3.1 of the cross product.*

Hint 103 *For property 7, consider the case when $\underline{a}$ and $\underline{b}$ are both non-zero. If they are parallel, use properties of 2×2 determinants in 4.1.4 to show that the cross product is zero. For the converse use the geometric interpretation (4.17) to show that $\underline{a} \times \underline{b} \neq \underline{0}$ when the vectors are not parallel.*

8. *Let $\underline{a}$, $\underline{b}$ and $\underline{c}$ be three vectors in $\Re^3$. Show that they are linearly independent if, and only if, the triple scalar product $\underline{a} \cdot (\underline{b} \times \underline{c})$ is not zero.*

9. **Consider the parallelogram $ABCD$ in Figure 4.2. The line EF passing through Q is*

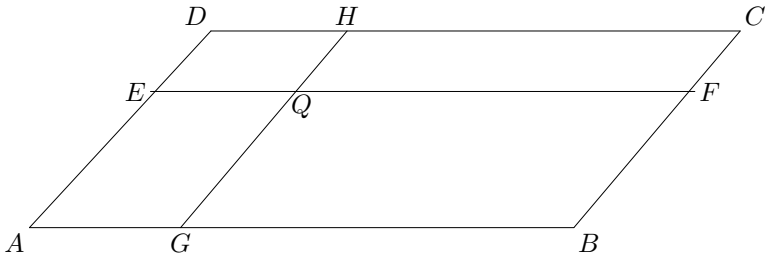


Figure 4.2

parallel to AB and the line GH passing through Q is parallel to AD. Using vector methods show that the parallelograms $EQHD$ and $GBFQ$ have equal areas if, and only if, Q lies on the diagonal AC. Conclude that this is the case if, and only if, the parallelograms $AGQE$ and $QFCH$ are similar. (A theorem going back to Euclid).

Hint 104 *Consider the cross product $\underline{AQ} \times \underline{QC}$.*

10. *A parallelepiped is a 'squashed box': it has six faces with opposite faces congruent parallelograms. Let $\underline{u} = \underline{AD}$, $\underline{v} = \underline{AB}$, and $\underline{w} = \underline{AE}$ be three of its linearly independent edges. Show that its volume is $|(\underline{u} \times \underline{v}) \cdot \underline{w}|$.*

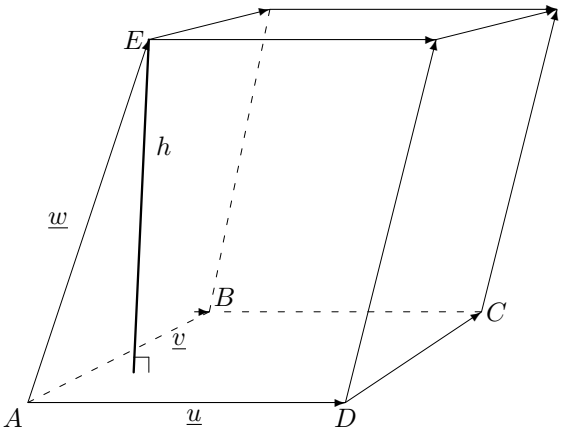


Figure 4.2 Parallelepiped

$h = \frac{\underline{w} \cdot (\underline{u} \times \underline{v})}{|\underline{u} \times \underline{v}|}$

Hint 105 *Let its base be $ABCD$. This has area $\alpha = |\underline{AD} \times \underline{AB}|$. The height h of the parallelepiped is the absolute value of the component of $\underline{AE}$ along $\underline{AD} \times \underline{AB}$. The volume is then $h\alpha$. Make a drawing.*

11. ** (The triple vector product). Let $\underline{a}$, $\underline{b}$ and $\underline{c}$ be three given vectors. Show that*

$$\underline{a} \times (\underline{b} \times \underline{c}) = (\underline{a} \cdot \underline{c})\, \underline{b} - (\underline{a} \cdot \underline{b})\, \underline{c}$$

The result is not altogether surprising since $\underline{a} \times (\underline{b} \times \underline{c})$ is a linear combination of $\underline{b}$ and $\underline{c}$.

Hint 106 *Show the result is true if $\underline{a}$ is $\underline{i}$, $\underline{j}$ or $\underline{k}$. Now use the result of No.2.*

12. ** Application to torques (moments) of forces.*
*Let $\underline{F}$ be a force vector that is applied to a point P in space. The **torque** (or **moment**) of $\underline{F}$ about the the point Q is defined as $\underline{M} = \underline{QP} \times \underline{F}$. The **line of action** of the force is the straight line ℓ parallel to $\underline{F}$ that passes through P. Show that the torques of $\underline{F}$ applied at any two points on its line of action about Q are equal.*
Let H be the point on ℓ closest to Q. Then with $\underline{F}$ acting at H we get the usual definition of 'moment', except that it now has a vector quality; its direction is perpendicular to both $\underline{QH}$ and $\underline{F}$.

13. ** (Continuation) Consider an axis ξ passing through Q and having the same direction as a given unit vector $\underline{l}$. Let D and E be points on the line of action of $\underline{F}$ and the axis ξ respectively such that the distance $|\underline{ED}|$ is minimum. (Thus $\underline{ED}$, if not zero, is perpendicular to both lines.) Let $\underline{m}$ be the unit vector with direction $\underline{ED}$, so that $\underline{ED} = \mu \underline{m}$, where $\mu = |\underline{ED}|$. Put $\underline{n} = \underline{l} \times \underline{m}$. In the sense of the right-hand rule, the torque of $\underline{F}$ about the **axis** ξ is the component $\underline{F} \cdot \underline{n}$ times the distance $\mu = |\underline{ED}|$ (draw a picture to see this).*
Show that the torque about the axis ξ is the triple scalar product $\underline{M} \cdot \underline{l}$.
Show further that if P' and Q' are any points on the line of action of $\underline{F}$ and the axis ξ respectively, then

$$\left(\underline{Q'P'} \times \underline{F}\right) \cdot \underline{l} = \left(\underline{QP} \times \underline{F}\right) \cdot \underline{l}$$

This shows that the torque of a force about an axis is purely a property of the force with its line of action and the axis and does not depend on selected points on these lines.

Hint 107 *Express $\underline{QD}$ as a linear combination $\lambda \underline{l} + \mu \underline{m}$ and $\underline{F}$ as a linear combination $\alpha \underline{l} + \gamma \underline{n}$. Now multiply out the expression $\underline{M} = (\lambda \underline{l} + \mu \underline{m}) \times (\alpha \underline{l} + \gamma \underline{n})$ and find $\underline{M} \cdot \underline{l}$.*

4.4 Summary of Chapter 4

4.4.1 Determinants

Determinants are inspired by the search for a formula to solve n equations for n unknowns. (See, for example, subsection 4.1.6).

4.4.2 2×2 determinants

In Exercise 3a of Chapter 2 we introduced the idea of the determinant: $|A| = a_{11}a_{22} - a_{21}a_{12}$. In subsection 4.1.4 of 4.1.1 we develop some basic properties of 2×2 determinants.

4.4.3 3×3 determinants

For a 3×3 matrix A, its determinant is given by equation (4.7):

$$|A| = \sum_{\sigma} \operatorname{sgn}(\sigma) a_{1\sigma 1} a_{2\sigma 2} a_{3\sigma 3}$$

Here the sum extends over all $\sigma = 3! = 6$ permutations of 1, 2, 3 and $\operatorname{sgn}(\sigma)$ is the sign of the permutation $\sigma = (\sigma 1, \sigma 2, \sigma 3)$.
A similar definition holds for an $n \times n$ matrix A.

4.4.4 Cofactors and properties of 3×3 determinants

(See subsections 4.1.10, 4.1.13, 4.1.14)The cofactor of $a_i j$ is -1^{i+j} times the determinant formed by deleting row i and column j.
A determinant can be expanded by any row:

$$|A| = a_{i1}A_{i|1} + a_{i2}A_{i|2} + a_{i3}A_{i|3} \tag{4.18}$$

(See equation (4.8).
Multiplying a row of A by a number k multiplies the determinant by k.
Interchanging two rows (columns) changes the sign of the determinant.
Adding a row (column) of A to another row (column) does not change the determinant.
To each property of rows there is a corresponding property of columns.

4.4.5 The adjoint

(See subsection 4.1.16)
The adjoint of A is given by

$$[\operatorname{adj}(A)]_{ij} = A_{j|i}$$

This means that we can first find the matrix B with $[B]_{ij} = A_{i|j}$ and then $\operatorname{adj}(A) = B^T$.

The fundamental property is

$$A\operatorname{adj}(A) = |A| I = \operatorname{adj}(A) A$$

Consequently, A has an inverse if, and only if, $|A| \neq 0$, in which case

$$A^{=1} = \frac{1}{|A|} \operatorname{adj}(A)$$

4.4.6 General determinants

All properties of 2×2 and 3×3 carry over to $n \times n$ determinants. (See section 4.2).

4.4.7 The Cross Product

(See section 4.3).

The **cross product** is defined in equation 4.14

$$\underline{a} \times \underline{b} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \underline{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \underline{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \underline{k}$$

4.4.8 The Triple Scalar Product of $\underline{a}$, $\underline{b}$ and $\underline{c}$

This is

$$\underline{a} \cdot (\underline{b} \times \underline{c})$$

Basic properties are given in equations (4.15) and (4.16).

4.4.9 Geometric interpretation

$$\underline{a} \times \underline{b} = (|\underline{a}||\underline{b}| \sin\theta)\,\widehat{\underline{c}}$$

Here $0 < \theta < \pi$ is the angle between the non-zero and non-parallel vectors $\underline{a}$ and $\underline{b}$ and $\widehat{\underline{c}}$ is the unit vector perpendicular to $\underline{a}$ and $\underline{b}$ with direction found by using the corkscrew rule: rotate from $\underline{a}$ to $\underline{b}$ through θ.
The cross product can also be seen geometrically as an area (exercise 3); the triple scalar product as a volume as a volume (exercise 10).

Chapter Eight: Theory of Space of the Vectors

2.1 Hilbert space.

2.1.1 Scalar products.

V is a complex vector space. A rule assigning to every pair of vectors $f, g \in V$ a complex number (f, g) is called a **semi-scalar product** if

1. (f, g) is linear in f when g is held fixed.
2. $(g, f) = \overline{(f, g)}$. This implies that (f, g) is anti-linear in g when f is held fixed. In other words. $(f, ag + bh) = \overline{a}(f, g) + \overline{b}(f, h)$. It also implies that (f, f) is real.
3. $(f, f) \geq 0$ for all $f \in V$.

If 3. is replaced by the stronger condition

4. $(f, f) > 0$ for all non-zero $f \in V$

then we say that $(\ ,\ )$ is a **scalar product**.

Examples.

- $V = \mathbf{C}^n$, so an element $\mathbf{z}$ of V is a column vector of complex numbers:

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

and $(\mathbf{z}, \mathbf{w})$ is given by

$$(\mathbf{z}, \mathbf{w}) := \sum_1^n z_i \overline{w_i}.$$

- V consists of all continuous (complex valued) functions on the real line which are periodic of period 2π and

$$(f,g) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

We will denote this space by $\mathcal{C}(\mathbf{T})$. Here the letter $\mathbf{T}$ stands for the one dimensional torus, i.e. the circle. We are identifying functions which are periodic with period 2π with functions which are defined on the circle $\mathbf{R}/2\pi\mathbf{Z}$.

- V consists of all doubly infinite sequences of complex numbers

$$\mathbf{a} = \ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots$$

which satisfy

$$\sum |a_i|^2 < \infty.$$

Here

$$(\mathbf{a}, \mathbf{b}) := \sum a_i \bar{b}_i.$$

All three are examples of scalar products.

2.1.2 The Cauchy-Schwartz inequality.

This says that if (,) is a semi-scalar product then

$$|(f,g)| \leq (f,f)^{\frac{1}{2}}(g,g)^{\frac{1}{2}}. \tag{2.1}$$

Proof. For any real number t condition 3. above says that $(f-tg, f-tg) \geq 0$. Expanding out gives

$$0 \leq (f-tg, f-tg) = (f,f) - t[(f,g)+(g,f)] + t^2(g,g).$$

Since $(g,f) = \overline{(f,g)}$, the coefficient of t in the above expression is twice the real part of (f,g). So the real quadratic form

$$Q(t) := (f,f) - 2\mathrm{Re}(f,g)t + t^2(g.g)$$

is nowhere negative. So it can not have distinct real roots, and hence by the $b^2 - 4ac$ rule we get

$$4(\mathrm{Re}(f,g))^2 - 4(f,f)(g,g) \leq 0$$

or

$$(\mathrm{Re}(f,g))^2 \leq (f,f)(g,g). \tag{2.2}$$

This is useful and almost but not quite what we want. But we may apply this inequality to $h = e^{i\theta}g$ for any θ. Then $(h,h) = (g,g)$. Choose θ so that

$$(f,g) = re^{i\theta}$$

where $r = |(f, g)|$. Then

$$(f, h) = (f, e^{i\theta} g) = e^{-i\theta}(f, g) = |(f, g)|$$

and the preceding inequality with g replaced by h gives

$$|(f, g)|^2 \leq (f, f)(g, g)$$

and taking square roots gives (2.1).

2.1.3 The triangle inequality

For any semiscalar product define

$$\|f\| := (f, f)^{\frac{1}{2}}$$

so we can write the Cauchy-Schwartz inequality as

$$|(f, g)| \leq \|f\|\|g\|.$$

The **triangle inequality** says that

$$\|f + g\| \leq \|f\| + \|g\|. \tag{2.3}$$

Proof.

$$\begin{aligned} \|f + g\|^2 &= (f + g, f + g) \\ &= (f, f) + 2\mathrm{Re}(f, g) + (g, g) \\ &\leq (f, f) + 2\|f\|\|g\| + (g, g) \quad \text{by (2.2)} \\ &= \|f|^2 + 2\|f\|\|g\| + \|g\|^2 \\ &= (\|f\| + \|g\|)^2. \end{aligned}$$

Taking square roots gives the triangle inequality (2.3). Notice that

$$\|cf\| = |c|\|f\| \tag{2.4}$$

since $(cf, cf) = c\overline{c}(f, f) = |c|^2\|f\|^2$.

Suppose we try to define the distance between two elements of V by

$$d(f, g) := \|f - g\|.$$

Notice that then $d(f, f) = 0$, $d(f, g) = d(g, f)$ and for any three elements

$$d(f, h) \leq d(f, g) + d(g, h)$$

by virtue of the triangle inequality. The only trouble with this definition is that we might have two distinct elements at zero distance, i.e. $0 = d(f, g) = \|f - g\|$. But this can not happen if $(\ ,\ )$ is a scalar product, i.e. satisfies condition 4.

A complex vector space V endowed with a scalar product is called a **pre-Hilbert** space.

Let V be a complex vector space and let $\|\cdot\|$ be a map which assigns to any $f \in V$ a non-negative real $\|f\|$ number such that $\|f\| > 0$ for all non-zero f. If $\|\cdot\|$ satisfies the triangle inequality (2.3) and equation (2.4) it is called a **norm**. A vector space endowed with a norm is called a normed space. The pre-Hilbert spaces can be characterized among all normed spaces by the parallelogram law as we will discuss below.

Later on, we will have to weaken condition (2.4) in our general study. But it is too complicated to give the general definition right now.

2.1.4 Hilbert and pre-Hilbert spaces.

The reason for the prefix "pre" is the following: The distance d defined above has all the desired properties we might expect of a distance. In particular, we can define the notions of "limit" and of a "Cauchy sequence" as is done for the real numbers: If f_n is a sequence of elements of V, and $f \in V$ we say that f is the limit of the f_n and write

$$\lim_{n\to\infty} f_n = f, \quad \text{or} \quad f_n \to f$$

if, for any positive number ϵ there is an $N = N(\epsilon)$ such that

$$d(f_n, f) < \epsilon \quad \text{for all } n \geq N.$$

If a sequence converges to some limit f, then this limit is unique, since any limits must be at zero distance and hence equal.

We say that a sequence of elements is **Cauchy** if for any $\delta > 0$ there is an $K = K(\delta)$ such that

$$d(f_m, f_n) < \delta \quad \forall, m, n \geq K.$$

If the sequence f_n has a limit, then it is Cauchy - just choose $K(\delta) = N(\frac{1}{2}\delta)$ and use the triangle inequality.

But it is quite possible that a Cauchy sequence has no limit. As an example of this type of phenomenon, think of the rational numbers with $|r - s|$ as the distance. The whole point of introducing the real numbers is to guarantee that every Cauchy sequence has a limit.

So we say that a pre-Hilbert space is a **Hilbert space** if it is "complete" in the above sense - if every Cauchy sequence has a limit.

Since the complex numbers are complete (because the real numbers are), it follows that $\mathbf{C}^n$ is complete, i.e. is a Hilbert space. Indeed, we can say that any finite dimensional pre-Hilbert space is a Hilbert space because it is isomorphic (as a pre-Hilbert space) to $\mathbf{C}^n$ for some n. (See below when we discuss orthonormal bases.)

The trouble is in the infinite dimensional case, such as the space of continuous periodic functions. This space is not complete. For example, let f_n be the function which is equal to one on $(-\pi + \frac{1}{n}, -\frac{1}{n})$, equal to zero on $(\frac{1}{n}, \pi - \frac{1}{n})$

and extended linearly $-\frac{1}{n}$ to $\frac{1}{n}$ and from $\pi - \frac{1}{n}$ to $\pi + \frac{1}{n}$ so as to be continuous and then extended so as to be periodic.(Thus on the interval $(\pi - \frac{1}{n}, \pi + \frac{1}{n})$ the function is given by $f_n(x) = \frac{n}{2}(x - (\pi - \frac{1}{n}).)$ If $m \leq n$, the functions f_m and f_n agree outside two intervals of length $\frac{2}{m}$ and on these intervals $|f_m(x) - f_n(x)| \leq 1$. So $\|f_m - f_n\|^2 \leq \frac{1}{2\pi} \cdot 2/m$ showing that the sequence $\{f_n\}$ is Cauchy. But the limit would have to equal one on $(-\pi, 0)$ and equal zero on $(0, \pi)$ and so be discontinuous at the origin and at π. Thus the space of continuous periodic functions is not a Hilbert space, only a pre-Hilbert space.

But just as we complete the rational numbers (by throwing in "ideal" elements) to get the real numbers, we may similarly complete any pre-Hilbert space to get a unique Hilbert space. See the section *Completion* in the chapter on metric spaces for a general discussion of how to complete any metric space. In particular, the completion of any normed vector space is a complete normed vector space. A complete normed space is called a **Banach** space. The general construction implies that any normed vector space can be completed to a Banach space. From the parallelogram law discussed below, it will follow that the completion of a pre-Hilbert space is a Hilbert space.

The completion of the space of continuous periodic functions will be denoted by $L_2(\mathbf{T})$.

2.1.5 The Pythagorean theorem.

Let V be a pre-Hilbert space. We have

$$\|f + g\|^2 = \|f\|^2 + 2\mathrm{Re}(f, g) + \|g\|^2.$$

So

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2 \quad \Leftrightarrow \ \mathrm{Re}\ (f, g) = 0. \tag{2.5}$$

We make the definition

$$f \perp g \ \Leftrightarrow \ (f, g) = 0$$

and say that f is perpendicular to g or that f is orthogonal to g. Notice that this is a stronger condition than the condition for the Pythagorean theorem, the right hand condition in (2.5). For example $\|f + if\|^2 = 2\|f\|^2$ but $(f, if) = -i\|f\|^2 \neq 0$ if $\|f\| \neq 0$.

If u_i is some finite collection of mutually orthogonal vectors, then so are $z_i u_i$ where the z_i are any complex numbers. So if

$$u = \sum_i z_i u_i$$

then by the Pythagorean theorem

$$\|u\|^2 = \sum_i |z_i|^2 \|u_i\|^2.$$

In particular, if the $u_i \neq 0$, then $u = 0 \Rightarrow z_i = 0$ for all i. This shows that any set of mutually orthogonal (non-zero) vectors is linearly independent.

Notice that the set of functions

$$e^{in\theta}$$

is an **orthonormal** set in the space of continuous periodic functions in that not only are they mutually orthogonal, but each has norm one.

2.1.6 The theorem of Apollonius.

Adding the equations

$$\begin{aligned} \|f+g\|^2 &= \|f\|^2 + 2\mathrm{Re}\ (f,g) + \|g\|^2 \quad (2.6) \\ \|f-g\|^2 &= \|f\|^2 - 2\mathrm{Re}\ (f,g) + \|g\|^2 \quad (2.7) \end{aligned}$$

gives

$$\|f+g\|^2 + \|f-g\|^2 = 2\left(\|f\|^2 + \|g\|^2\right). \quad (2.8)$$

This is known as the **parallelogram law**. It is the algebraic expression of the theorem of Apollonius which asserts that the sum of the areas of the squares on the sides of a parallelogram equals the sum of the areas of the squares on the diagonals.

If we subtract (2.7) from (2.6) we get

$$\mathrm{Re}(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2\right). \quad (2.9)$$

Now $(if,g) = i(f,g)$ and $\mathrm{Re}\{i(f,g)\} = -\mathrm{Im}(f,g)$ so

$$\mathrm{Im}(f,g) = -\mathrm{Re}(if,g) = \mathrm{Re}(f,ig)$$

so

$$(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2 + i\|f+ig\|^2 - i\|f-ig\|^2\right). \quad (2.10)$$

If we now complete a pre-Hilbert space, the right hand side of this equation is defined on the completion, and is a continuous function there. It therefore follows that the scalar product extends to the completion, and, by continuity, satisfies all the axioms for a scalar product, plus the completeness condition for the associated norm. In other words, the completion of a pre-Hilbert space is a Hilbert space.

2.1.7 The theorem of Jordan and von Neumann.

This is essentially a converse to the theorem of Apollonius. It says that if $\|\cdot\|$ is a norm on a (complex) vector space V which satisfies (2.8), then V is in fact a pre-Hilbert space with $\|f\|^2 = (f,f)$. If the theorem is true, then the scalar product must be given by (2.10). So we must prove that if we take (2.10) as the

definition, then all the axioms on a scalar product hold. The easiest axiom to verify is

$$(g, f) = \overline{(f, g)}.$$

Indeed, the real part of the right hand side of (2.10) is unchanged under the interchange of f and g (since $g - f = -(f - g)$ and $\|-h\| = \|h\|$ for any h is one of the properties of a norm). Also $g + if = i(f - ig)$ and $\|ih\| = \|h\|$ so the last two terms on the right of (2.10) get interchanged, proving that $(g, f) = \overline{(f, g)}$.

It is just as easy to prove that

$$(if, g) = i(f, g).$$

Indeed replacing f by if sends $\|f + ig\|^2$ into $\|if + ig\|^2 = \|f + g\|^2$ and sends $\|f + g\|^2$ into $\|if + g\|^2 = \|i(f - ig)\|^2 = \|f - ig\|^2 = i(-i\|f - ig\|^2)$ so has the effect of multiplying the sum of the first and fourth terms by i, and similarly for the sum of the second and third terms on the right hand side of (2.10).

Now (2.10) implies (2.9). Suppose we replace f, g in (2.8) by $f_1 + g, f_2$ and by $f_1 - g, f_2$ and subtract the second equation from the first. We get

$$\|f_1 + f_2 + g\|^2 - \|f_1 + f_2 - g\|^2 + \|f_1 - f_2 + g\|^2 - \|f_1 - f_2 - g\|^2$$

$$= 2\left(\|f_1 + g\|^2 - \|f_1 - g\|^2\right).$$

In view of (2.9) we can write this as

$$\text{Re}\,(f_1 + f_2, g) + \text{Re}\,(f_1 - f_2, g) = 2\text{Re}\,(f_1, g). \tag{2.11}$$

Now the right hand side of (2.9) vanishes when $f = 0$ since $\|g\| = \| - g\|$. So if we take $f_1 = f_2 = f$ in (2.11) we get

$$\text{Re}\,(2f, g) = 2\text{Re}\,(f, g).$$

We can thus write (2.11) as

$$\text{Re}\,(f_1 + f_2, g) + \text{Re}\,(f_1 - f_2, g) = \text{Re}\,(2f_1, g).$$

In this equation make the substitutions

$$f_1 \mapsto \frac{1}{2}(f_1 + f_2), \quad f_2 \mapsto \frac{1}{2}(f_1 - f_2).$$

This yields

$$\text{Re}\,(f_1 + f_2, g) = \text{Re}\,(f_1, g) + \text{Re}\,(f_2, g).$$

Since it follows from (2.10) and (2.9) that

$$(f, g) = \text{Re}\,(f, g) - i\text{Re}\,(if, g)$$

we conclude that

$$(f_1 + f_2, g) = (f_1, g) + (f_2, g).$$

Taking $f_1 = -f_2$ shows that

$$(-f, g) = -(f, g).$$

Consider the collection $\mathcal{C}$ of complex numbers α which satisfy

$$(\alpha f, g) = \alpha(f, g)$$

(for all f, g). We know from $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$ that

$$\alpha, \beta \in \mathcal{C} \Rightarrow \alpha + \beta \in \mathcal{C}.$$

So $\mathcal{C}$ contains all integers. If $0 \neq \beta \in \mathcal{C}$ then

$$(f, g) = (\beta \cdot (1/\beta)f, g) = \beta((1/\beta)f, g)$$

so $\beta^{-1} \in \mathcal{C}$. Thus $\mathcal{C}$ contains all (complex) rational numbers. The theorem will be proved if we can prove that $(\alpha f, g)$ is continuous in α. But the triangle inequality

$$\|f + g\| \leq \|f\| + \|g\|$$

applied to $f = f_2, g = f_1 - f_2$ implies that

$$\|f_1\| \leq \|f_1 - f_2\| + \|f_2\|$$

or

$$\|f_1\| - \|f_2\| \leq \|f_1 - f_2\|.$$

Interchanging the role of f_1 and f_2 gives

$$| \, \|f_1\| - \|f_2\| \, | \leq \|f_1 - f_2\|.$$

Therefore

$$| \, \|\alpha f \pm g\| - \|\beta f \pm g\| \, | \leq \|(\alpha - \beta)f\|.$$

Since $\|(\alpha - \beta)f\| \to 0$ as $\alpha \to \beta$ this shows that the right hand side of (2.10) when applied to αf and g is a continuous function of α. Thus $\mathcal{C} = \mathbf{C}$. We have proved

Theorem 2.1.1 [P. Jordan and J. von Neumann] *If V is a normed space whose norm satisfies (2.8) then V is a pre-Hilbert space.*

Notice that the condition (2.8) involves only two vectors at a time. So we conclude as an immediate consequence of this theorem that

Corollary 2.1.1 *A normed vector space is pre-Hilbert space if and only if every two dimensional subspace is a Hilbert space in the induced norm.*

Actually, a weaker version of this corollary, with two replaced by three had been proved by Fréchet, *Annals of Mathematics*, July 1935, who raised the problem of giving an abstract characterization of those norms on vector spaces which come from scalar products. In the immediately following paper Jordan and von Neumann proved the theorem above leading to the stronger corollary that two dimensions suffice.

2.1.8 Orthogonal projection.

We continue with the assumption that V is pre-Hilbert space. If A and B are two subsets of V, we write $A \perp B$ if $u \in A$ and $v \in B \Rightarrow u \perp v$, in other words if every element of A is perpendicular to every element of B. Similarly, we will write $v \perp A$ if the element v is perpendicular to all elements of A. Finally, we will write $A^\perp$ for the set of all v which satisfy $v \perp A$. Notice that $A^\perp$ is always a linear subspace of V, for any A.

Now let M be a (linear) subspace of V. Let v be some element of V, not necessarily belonging to M. We want to investigate the problem of finding a $w \in M$ such that $(v - w) \perp M$. Of course, if $v \in M$ then the only choice is to take $w = v$. So the interesting problem is when $v \notin M$. Suppose that such a w exists, and let x be any (other) point of M. Then by the Pythagorean theorem,

$$\|v - x\|^2 = \|(v - w) + (w - x)\|^2 = \|v - w\|^2 + \|w - x\|^2$$

since $(v - w) \perp M$ and $(w - x) \in M$. So

$$\|v - w\| \leq \|v - x\|$$

and this inequality is strict if $x \neq w$. In words: if we can find a $w \in M$ such that $(v - w) \perp M$ then w is the unique solution of the problem of finding the point in M which is closest to v. Conversely, suppose we found a $w \in M$ which has this minimization property, and let x be any element of M. Then for any real number t we have

$$\|v - w\|^2 \leq \|(v - w) + tx\|^2 = \|v - w\|^2 + 2t\text{Re } (v - w, x) + t^2\|x\|^2.$$

Since the minimum of this quadratic polynomial in t occurring on the right is achieved at $t = 0$, we conclude (by differentiating with respect to t and setting $t = 0$, for example) that

$$\text{Re } (v - w, x) = 0.$$

By our usual trick of replacing x by $e^{i\theta}x$ we conclude that

$$(v - w, x) = 0.$$

Since this holds for all $x \in M$, we conclude that $(v - w) \perp M$. So to find w we search for the minimum of $\|v - x\|,\ x \in M$.

Now $\|v - x\| \geq 0$ and is some finite number for any $x \in M$. So there will be some real number m such that $m \leq \|v - x\|$ for $x \in M$, and such that no strictly larger real number will have this property. (m is known as the "greatest lower bound" of the values $\|v - x\|,\ x \in M$.) So we can find a sequence of vectors $x_n \in M$ such that

$$\|v - x_n\| \to m.$$

We claim that the x_n form a Cauchy sequence. Indeed,

$$\|x_i - x_j\|^2 = \|(v - x_j) - (v - x_i)\|^2$$

and by the parallelogram law this equals

$$2\left(\|v-x_i\|^2+\|v-x_j\|^2\right)-\|2v-(x_i+x_j)\|^2.$$

Now the expression in parenthesis converges to $2m^2$. The last term on the right is

$$-\|2(v-\frac{1}{2}(x_i+x_j)\|^2.$$

Since $\frac{1}{2}(x_i+x_j)\in M$, we conclude that

$$\|2v-(x_i+x_j)\|^2\geq 4m^2$$

so

$$\|x_i-x_j\|^2\leq 4(m+\epsilon)^2-4m^2$$

for i and j large enough that $\|v-x_i\|\leq m+\epsilon$ and $\|v-x_j\|\leq m+\epsilon$. This proves that the sequence x_n is Cauchy.

Here is the crux of the matter: If M is complete, then we can conclude that the x_n converge to a limit w which is then the unique element in M such that $(v-w)\perp M$. It is at this point that completeness plays such an important role.

Put another way, we can say that if M is a subspace of V which is complete (under the scalar product $(\ ,\ )$ restricted to M) then we have the orthogonal direct sum decomposition

$$V=M\oplus M^{\perp},$$

which says that every element of V can be uniquely decomposed into the sum of an element of M and a vector perpendicular to M.

For example, if M is the one dimensional subspace consisting of all (complex) multiples of a non-zero vector y, then M is complete, since $\mathbf{C}$ is complete. So w exists. Since all elements of M are of the form ay, we can write $w=ay$ for some complex number a. Then $(v-ay,y)=0$ or

$$(v,y)=a\|y\|^2$$

so

$$a=\frac{(v,y)}{\|y\|^2}.$$

We call a the **Fourier coefficient** of v with respect to y. Particularly useful is the case where $\|y\|=1$ and we can write

$$a=(v,y). \tag{2.12}$$

Getting back to the general case, if $V=M\oplus M^{\perp}$ holds, so that to every v there corresponds a unique $w\in M$ satisfying $(v-w)\in M^{\perp}$ the map $v\mapsto w$ is called orthogonal projection of V onto M and will be denoted by π_M.

2.1.9 The Riesz representation theorem.

Let V and W be two complex vector spaces. A map

$$T : V \to W$$

is called **linear** if

$$T(\lambda x + \mu y) = \lambda T(x) + \mu T(Y) \quad \forall\ x, y \in V, \quad \lambda, \mu \in \mathbf{C}$$

and is called **anti-linear** if

$$T(\lambda x + \mu y) = \overline{\lambda} T(x) + \overline{\mu} T(Y) \quad \forall\ x, y \in V \quad \lambda, \mu \in \mathbf{C}.$$

If $\ell : V \to \mathbf{C}$ is a linear map, (also known as a linear function) then

$$\ker\ \ell := \{x \in V |\ \ell(x) = 0\}$$

has codimension one (unless $\ell \equiv 0$). Indeed, if

$$\ell(y) \neq 0$$

then

$$\ell(x) = 1 \ \text{ where } x = \frac{1}{\ell(y)} y$$

and for any $z \in V$,

$$z - \ell(z)x \in \ker\ \ell.$$

If V is a normed space and ℓ is continuous, then ker (ℓ) is a closed subspace. The space of continuous linear functions is denoted by V^*. It has its own norm defined by

$$\|\ell\| := \sup_{x \in V, \|x\| \neq 0} |\ell(x)| / \|x\|.$$

Suppose that H is a pre-hilbert space. Then we have an antilinear map

$$\phi : H \to H^*, \quad (\phi(g))(f) := (f, g).$$

The Cauchy-Schwartz inequality implies that

$$\|\phi(g)\| \leq \|g\|$$

and in fact

$$(g, g) = \|g\|^2$$

shows that

$$\|\phi(g)\| = \|g\|.$$

In particular the map ϕ is injective.

The Riesz representation theorem says that if H is a Hilbert space, then this map is surjective:

Theorem 2.1.2 *Every continuous linear function on H is given by scalar product by some element of H.*

The proof is a consequence of the theorem about projections applied to

$$N := \ker \ell :$$

If $\ell = 0$ there is nothing to prove. If $\ell \neq 0$ then N is a closed subspace of codimension one. Choose $v \notin N$. Then there is an $x \in N$ with $(v - x) \perp N$. Let

$$y := \frac{1}{\|v - x\|}(v - x).$$

Then

$$y \perp N$$

and

$$\|y\| = 1.$$

For any $f \in H$,

$$[f - (f, y)y] \perp y$$

so

$$f - (f, y)y \in N$$

or

$$\ell(f) = (f, y)\ell(y),$$

so if we set

$$g := \overline{\ell(y)}y$$

then

$$(f, g) = \ell(f)$$

for all $f \in H$. QED

2.1.10 What is $L_2(\mathbf{T})$?

We have defined the space $L^2(\mathbf{T})$ to be the completion of the space $\mathcal{C}(\mathbf{T})$ under the L_2 norm $\|f\|_2 = (f, f)^{\frac{1}{2}}$. In particular, every linear function on $\mathcal{C}(\mathbf{T})$ which is continuous with respect to the this L_2 norm extends to a unique continuous linear function on $L_2(\mathbf{T})$. By the Riesz representation theorem we know that every such continuous linear function is given by scalar product by an element of $L_2(\mathbf{T})$. Thus we may think of the elements of $L_2(\mathbf{T})$ as being the linear functions on $\mathcal{C}(\mathbf{T})$ which are continuous with respect to the L_2 norm. An element of $L_2(\mathbf{T})$ should not be thought of as a function, but rather as a linear function on the space of continuous functions relative to a special norm - the L_2 norm.

2.1.11 Projection onto a direct sum.

Suppose that the closed subspace M of a pre-Hilbert space is the orthogonal direct sum of a finite number of subspaces

$$M = \bigoplus_i M_i$$

meaning that the M_i are mutually perpendicular and every element x of M can be written as

$$x = \sum x_i, \quad x_i \in M_i.$$

(The orthogonality guarantees that such a decomposition is unique.) Suppose further that each M_i is such that the projection π_{M_i} exists. Then π_M exists and

$$\pi_M(v) = \sum \pi_{M_i}(v). \tag{2.13}$$

Proof. Clearly the right hand side belongs to M. We must show $v - \sum_i \pi_{M_i}(v)$ is orthogonal to every element of M. For this it is enough to show that it is orthogonal to each M_j since every element of M is a sum of elements of the M_j. So suppose $x_j \in M_j$. But $(\pi_{M_i} v, x_j) = 0$ if $i \neq j$. So

$$(v - \sum \pi_{M_i}(v), x_j) = (v - \pi_{M_j}(v), x_j) = 0$$

by the defining property of π_{M_j}.

2.1.12 Projection onto a finite dimensional subspace.

We now will put the equations (2.12) and (2.13) together: Suppose that M is a finite dimensional subspace with an orthonormal basis ϕ_i. This implies that M is an orthogonal direct sum of the one dimensional spaces spanned by the ϕ_i and hence π_M exists and is given by

$$\pi_M(v) = \sum a_i \phi_i \quad \text{where} \quad a_i = (v, \phi_i). \tag{2.14}$$

2.1.13 Bessel's inequality.

We now look at the infinite dimensional situation and suppose that we are given an orthonormal sequence $\{\phi_i\}_1^\infty$. Any $v \in V$ has its Fourier coefficients

$$a_i = (v, \phi_i)$$

relative to the members of this sequence. Bessel's inequality asserts that

$$\sum_1^\infty |a_i|^2 \leq \|v\|^2, \tag{2.15}$$

in particular the sum on the left converges.

Proof. Let

$$v_n := \sum_{i=1}^{n} a_i \phi_i,$$

so that v_n is the projection of v onto the subspace spanned by the first n of the ϕ_i. In any event, $(v - v_n) \perp v_n$ so by the Pythagorean Theorem

$$\|v\|^2 = \|v - v_n\|^2 + \|v_n\|^2 = \|v - v_n\|^2 + \sum_{i=1}^{n} |a_i|^2.$$

This implies that

$$\sum_{i=1}^{n} |a_i|^2 \leq \|v\|^2$$

and letting $n \to \infty$ shows that the series on the left of Bessel's inequality converges and that Bessel's inequality holds.

2.1.14 Parseval's equation.

Continuing the above argument, observe that

$$\|v - v_n\|^2 \to 0 \;\Leftrightarrow\; \sum |a_i|^2 = \|v\|^2.$$

But $\|v - v_n\|^2 \to 0$ if and only if $\|v - v_n\| \to 0$ which is the same as saying that $v_n \to v$. But v_n is the n-th partial sum of the series $\sum a_i \phi_i$, and in the language of series, we say that a series converges to a limit v and write $\sum a_i \phi_i = v$ if and only if the partial sums approach v. So

$$\sum a_i \phi_i = v \quad \Leftrightarrow \quad \sum_i |a_i|^2 = \|v\|^2. \tag{2.16}$$

In general, we will call the series $\sum_i a_i \phi_i$ the Fourier series of v (relative to the given orthonormal sequence) whether or not it converges to v. Thus Parseval's equality says that the Fourier series of v converges to v if and only if $\sum |a_i|^2 = \|v\|^2$.

2.1.15 Orthonormal bases.

We still suppose that V is merely a pre-Hilbert space. We say that an orthonormal sequence $\{\phi_i\}$ is a **basis** of V if every element of V is the sum of its Fourier series. For example, one of our tasks will be to show that the exponentials $\{e^{inx}\}_{n=-\infty}^{\infty}$ form a basis of $\mathcal{C}(\mathbf{T})$.

If the orthonormal sequence ϕ_i is a basis, then any v can be approximated as closely as we like by finite linear combinations of the ϕ_i, in fact by the partial sums of its Fourier series. We say that the finite linear combinations of the ϕ_i are *dense* in V. Conversely, suppose that the finite linear combinations of the

ϕ_i are dense in V. This means that for any v and any $\epsilon > 0$ we can find an n and a set of n complex numbers b_i such that

$$\|v - \sum b_i \phi_i\| \leq \epsilon.$$

But we know that v_n is the closest vector to v among all the linear combinations of the first n of the ϕ_i. so we must have

$$\|v - v_n\| \leq \epsilon.$$

But this says that the Fourier series of v converges to v, i.e. that the ϕ_i form a basis. For example, we know from Fejer's theorem that the exponentials e^{ikx} are dense in $\mathcal{C}(\mathbf{T})$. Hence we know that they form a basis of the pre-Hilbert space $\mathcal{C}(\mathbf{T})$. We will give some alternative proofs of this fact below.

In the case that V is actually a Hilbert space, and not merely a pre-Hilbert space, there is an alternative and very useful criterion for an orthonormal sequence to be a basis: Let M be the set of all limits of finite linear combinations of the ϕ_i. Any Cauchy sequence in M converges (in V) since V is a Hilbert space, and this limit belongs to M since it is itself a limit of finite linear combinations of the ϕ_i (by the diagonal argument for example). Thus $V = M \oplus M^{\perp}$, and the ϕ_i form a basis of M. So the ϕ_i form a basis of V if and only if $M^{\perp} = \{0\}$. But this is the same as saying that no non-zero vector is orthogonal to all the ϕ_i. So we have proved

Proposition 2.1.1 *In a Hilbert space, the orthonormal set $\{\phi_i\}$ is a basis if and only if no non-zero vector is orthogonal to all the ϕ_i.*

2.2 Self-adjoint transformations.

We continue to let V denote a pre-Hilbert space. Let T be a linear transformation of V into itself. This means that for every $v \in V$ the vector $Tv \in V$ is defined and that Tv depends linearly on v : $T(av + bw) = aTv + bTw$ for any two vectors v and w and any two complex numbers a and b. We recall from linear algebra that a non-zero vector v is called an eigenvector of T if Tv is a scalar times v, in other words if $Tv = \lambda v$ where the number λ is called the corresponding eigenvalue.

A linear transformation T on V is called **symmetric** if for any pair of elements v and w of V we have

$$(Tv, w) = (v, Tw).$$

Notice that if v is an eigenvector of a symmetric transformation T with eigenvalue λ, then

$$\lambda(v, v) = (\lambda v, v) = (Tv, v) = (v, Tv) = (v, \lambda v) = \overline{\lambda}(v, v),$$

so $\lambda = \overline{\lambda}$. In other words, all eigenvalues of a symmetric transformation are real.

We will let $\mathbf{S} = \mathbf{S}(V)$ denote the "unit sphere" of V, i.e. $\mathbf{S}$ denotes the set of all $\phi \in V$ such that $\|\phi\| = 1$. A linear transformation T is called **bounded** if $\|T\phi\|$ is bounded as ϕ ranges over all of $\mathbf{S}$. If T is bounded, we let

$$\|T\| := \max_{\phi \in \mathbf{S}} \|T\phi\|.$$

Then

$$\|Tv\| \leq \|T\|\|v\|$$

for all $v \in V$. A linear transformation on a finite dimensional space is automatically bounded, but not so for an infinite dimensional space.

Also, for any linear transformation T, we will let $N(T)$ denote the kernel of T, so

$$N(T) = \{v \in V | Tv = 0\}$$

and $R(T)$ denote the range of T, so

$$R(T) := \{v | v = Tw \text{ for some } w \in V\}.$$

Both $N(T)$ and $R(T)$ are linear subspaces of V.

For bounded transformations, the phrase "self-adjoint" is synonymous with "symmetric". Later on we will need to study non-bounded (not everywhere defined) symmetric transformations, and then a rather subtle and important distinction will be made between self-adjoint transformations and those which are merely symmetric. But for the rest of this section we will only be considering bounded linear transformations, and so we will freely use the phrase "self-adjoint", and (usually) drop the adjective "bounded" since all our transformations will be assumed to be bounded.

We denote the set of all (bounded) self-adjoint transformations by $\mathcal{A}$, or by $\mathcal{A}(V)$ if we need to make V explicit.

2.2.1 Non-negative self-adjoint transformations.

If T is a self-adjoint transformation, then

$$(Tv, v) = (v, Tv) = \overline{(Tv, v)}$$

so (Tv, v) is always a real number. More generally, for any pair of elements v and w,

$$(Tv, w) = \overline{(Tw, v)}.$$

Since (Tv, w) depends linearly on v for fixed w, we see that the rule which assigns to every pair of elements v and w the number (Tv, w) satisfies the first two conditions in our definition of a semi-scalar product. Since (Tv, v) might be negative, condition 3. of the definition need not be satisfied. This leads to the following definition:

A self-adjoint transformation T is called **non-negative** if

$$(Tv, v) \geq 0 \quad \forall\ v \in V.$$

So if T is a non-negative self-adjoint transformation, then the rule which assigns to every pair of elements v and w the number (Tv, w) *is* a semi-scalar product to which we may apply the Cauchy-Schwartz inequality and conclude that

$$|(Tv, w)| \leq (Tv, v)^{\frac{1}{2}}(Tw, w)^{\frac{1}{2}}.$$

Now let us assume in addition that T is bounded with norm $\|T\|$. Let us take $w = Tv$ in the preceding inequality. We get

$$\|Tv\|^2 = |(Tv, Tv)| \leq (Tv, v)^{\frac{1}{2}}(TTv, Tv)^{\frac{1}{2}}.$$

Now apply the Cauchy-Schwartz inequality for the original scalar product to the last factor on the right:

$$(TTv, Tv)^{\frac{1}{2}} \leq \|TTv\|^{\frac{1}{2}}\|Tv\|^{\frac{1}{2}} \leq \|T\|^{\frac{1}{2}}\|Tv\|^{\frac{1}{2}}\|Tv\|^{\frac{1}{2}} = \|T\|^{\frac{1}{2}}\|Tv\|,$$

where we have used the defining property of $\|T\|$ in the form $\|TTv\| \leq \|T\|\|Tv\|$. Substituting this into the previous inequality we get

$$\|Tv\|^2 \leq (Tv, v)^{\frac{1}{2}}\|T\|^{\frac{1}{2}}\|Tv\|.$$

If $\|Tv\| \neq 0$ we may divide this inequality by $\|Tv\|$ to obtain

$$\|Tv\| \leq \|T\|^{\frac{1}{2}}(Tv, v)^{\frac{1}{2}}. \tag{2.17}$$

This inequality is clearly true if $\|Tv\| = 0$ and so holds in all cases.

We will make much use of this inequality. For example, it follows from (2.17) that

$$(Tv, v) = 0 \quad \Rightarrow \quad Tv = 0. \tag{2.18}$$

It also follows from (2.17) that if we have a sequence $\{v_n\}$ of vectors with $(Tv_n, v_n) \to 0$ then $\|Tv_v\| \to 0$ and so

$$(Tv_n, v_n) \to 0 \quad \Rightarrow \quad Tv_n \to 0. \tag{2.19}$$

Notice that if T is a bounded self adjoint transformation, not necessarily non-negative, then $rI - T$ is a non-negative self-adjoint transformation if $r \geq \|T\|$: Indeed,

$$((rI - T)v, v) = r(v, v) - (Tv, v) \geq (r - \|T\|)(v, v) \geq 0$$

since, by Cauchy-Schwartz,

$$(Tv, v) \leq |(Tv, v)| \leq \|Tv\|\|v\| \leq \|T\|\|v\|^2 = \|T\|(v, v).$$

So we may apply the preceding results to $rI - T$.

2.3 Compact self-adjoint transformations.

We say that the self-adjoint transformation T is **compact** if it has the following property: Given any sequence of elements $u_n \in \mathbf{S}$, we can choose a subsequence u_{n_i} such that the sequence Tu_{n_i} converges to a limit in V.

Some remarks about this complicated looking definition: In case V is finite dimensional, every linear transformation is bounded, hence the sequence Tu_n lies in a bounded region of our finite dimensional space, and hence by the completeness property of the real (and hence complex) numbers, we can always find such a convergent subsequence. So in finite dimensions every T is compact. More generally, the same argument shows that if $R(T)$ is finite dimensional and T is bounded then T is compact. So the definition is of interest essentially in the case when $R(T)$ is infinite dimensional.

Also notice that if T is compact, then T is bounded. Otherwise we could find a sequence u_n of elements of $\mathbf{S}$ such that $\|Tu_n\| \geq n$ and so no subsequence Tu_{n_i} can converge.

We now come to the key result which we will use over and over again:

Theorem 2.3.1 *Let T be a compact self-adjoint operator. Then $R(T)$ has an orthonormal basis $\{\phi_i\}$ consisting of eigenvectors of T and if $R(T)$ is infinite dimensional then the corresponding sequence $\{r_n\}$ of eigenvalues converges to* 0.

Proof. We know that T is bounded. If $T = 0$ there is nothing to prove. So assume that $T \neq 0$ and let

$$m_1 := \|T\| > 0.$$

By the definition of $\|T\|$ we can find a sequence of vectors $u_n \in \mathbf{S}$ such that $\|Tu_n\| \to \|T\|$. By the definition of compactness we can find a subsequence of this sequence so that $Tu_{n_i} \to w$ for some $w \in V$. On the other hand, the transformation T^2 is self-adjoint and bounded by $\|T\|^2$. Hence $\|T\|^2 I - T^2$ is non-negative, and

$$((\|T\|^2 I - T^2)u_n, u_n) = \|T\|^2 - \|Tu_n\|^2 \to 0.$$

So we know from (2.19) that

$$\|T\|^2 u_n - T^2 u_n \to 0.$$

Passing to the subsequence we have $T^2 u_{n_i} = T(Tu_{n_i}) \to Tw$ and so

$$\|T\|^2 u_{n_i} \to Tw$$

or

$$u_{n_i} \to \frac{1}{m_1^2} Tw.$$

Applying T to this we get

$$Tu_{n_i} \to \frac{1}{m_1^2} T^2 w$$

or

$$T^2 w = m_1^2 w.$$

Also $\|w\| = \|T\| = m_1 \neq 0$. So $w \neq 0$. So w is an eigenvector of T^2 with eigenvalue m_1^2. We have

$$0 = (T^2 - m_1^2)w = (T + m_1)(T - m_1)w.$$

If $(T - m_1)w = 0$, then w is an eigenvector of T with eigenvalue m_1 and we normalize by setting

$$\phi_1 := \frac{1}{\|w\|} w.$$

Then $\|\phi_1\| = 1$ and

$$T\phi_1 = m_1 \phi_1.$$

If $(T - m_1)w \neq 0$ then $y := (T - m_1)w$ is an eigenvector of T with eigenvalue $-m_1$ and again we normalize by setting

$$\phi_1 := \frac{1}{\|y\|} y.$$

So we have found a unit vector $\phi_1 \in R(T)$ which is an eigenvector of T with eigenvalue $r_1 = \pm m_1$.

Now let

$$V_2 := \phi_1^{\perp}.$$

If $(w, \phi_1) = 0$, then

$$(Tw, \phi_1) = (w, T\phi_1) = r_1(w, \phi_1) = 0.$$

In other words,

$$T(V_2) \subset V_2$$

and we can consider the linear transformation T restricted to V_2 which is again compact. If we let m_2 denote the norm of the linear transformation T when restricted to V_2 then $m_2 \leq m_1$ and we can apply the preceding procedure to find a unit eigenvector ϕ_2 with eigenvalue $\pm m_2$.

We proceed inductively, letting

$$V_n := \{\phi_1, \ldots, \phi_{n-1}\}^{\perp}$$

and find an eigenvector ϕ_n of T restricted to V_n with eigenvalue $\pm m_n \neq 0$ if the restriction of T to V_n is not zero. So there are two alternatives:

- after some finite stage the restriction of T to V_n is zero. In this case $R(T)$ is finite dimensional with orthonormal basis $\phi_1, \ldots, \phi_{n-1}$. Or
- The process continues indefinitely so that at each stage the restriction of T to V_n is not zero and we get an infinite sequence of eigenvectors and eigenvalues r_i with $|r_i| \geq |r_{i+1}|$.

The first case is one of the alternatives in the theorem, so we need to look at the second alternative.

We first prove that $|r_n| \to 0$. If not, there is some $c > 0$ such that $|r_n| \geq c$ for all n (since the $|r_n|$ are decreasing). If $i \neq j$,then by the Pythagorean theorem we have

$$\|T\phi_i - T\phi_j\|^2 = \|r_i\phi_i - r_j\phi_j\|^2 = r_i^2\|\phi_i\|^2 + r_j^2\|\phi_j\|^2.$$

Since $\|\phi_i\| = \|\phi_j| = 1$ this gives

$$\|T\phi_i - T\phi_j\|^2 = r_i^2 + r_j^2 \geq 2c^2.$$

Hence no subsequence of the $T\phi_i$ can converge, since all these vectors are at least a distance $c\sqrt{2}$ apart. This contradicts the compactness of T.

To complete the proof of the theorem we must show that the ϕ_i form a basis of $R(T)$. So if $w = Tv$ we must show that the Fourier series of w with respect to the ϕ_i converges to w. We begin with the Fourier coefficients of v relative to the ϕ_i which are given by

$$a_n = (v, \phi_n).$$

Then the Fourier coefficients of w are given by

$$b_i = (w, \phi_i) = (Tv, \phi_i) = (v, T\phi_i) = (v, r_i\phi_i) = r_i a_i.$$

So

$$w - \sum_{i=1}^{n} b_i\phi_i = Tv - \sum_{i=1}^{n} a_i r_i \phi_i = T(v - \sum_{i=1}^{n} a_i\phi_i).$$

Now $v - \sum_{i=1}^{n} a_i\phi_i$ is orthogonal to $\phi_1, \ldots, \phi_n$ and hence belongs to V_{n+1}. So

$$\|T(v - \sum_{i=1}^{n} a_i\phi_i)\| \leq |r_{n+1}|\|(v - \sum_{i=1}^{n} a_i\phi_i)\|.$$

By the Pythagorean theorem,

$$\|(v - \sum_{i=1}^{n} a_i\phi_i)\| \leq \|v\|.$$

Putting the two previous inequalities together we get

$$\|w - \sum_{i=1}^{n} b_i\phi_i\| = \|T(v - \sum_{i=1}^{n} a_i\phi_i)\| \leq |r_{n+1}|\|v\| \to 0.$$

This proves that the Fourier series of w converges to w concluding the proof of the theorem.

The "converse" of the above result is easy. Here is a version: Suppose that $\mathbf{H}$ is a Hilbert space with an orthonormal basis $\{\phi_i\}$ consisting of eigenvectors

of an operator T, so $T\phi_i = \lambda_i \phi_i$, and suppose that $\lambda_i \to 0$ as $i \to \infty$. Then T is compact. Indeed, for each j we can find an $N = N(j)$ such that

$$|\lambda_r| < \frac{1}{j} \quad \forall\ r > N(j).$$

We can then let $\mathbf{H}_j$ denote the closed subspace spanned by all the eigenvectors $\phi_r, r > N(j)$, so that

$$\mathbf{H} = \mathbf{H}_j^{\perp} \oplus \mathbf{H}_j$$

is an orthogonal decomposition and $\mathbf{H}_j^{\perp}$ is finite dimensional, in fact is spanned the first $N(j)$ eigenvectors of T.

Now let $\{u_i\}$ be a sequence of vectors with $\|u_i\| \leq 1$ say. We decompose each element as

$$u_i = u_i' \oplus u_i'', \qquad u_i' \in \mathbf{H}_1^{\perp},\ u_i'' \in \mathbf{H}_j.$$

We can choose a subsequence so that u_{i_k}' converges, because they all belong to a finite dimensional space, and hence so does Tu_{i_k} since T is bounded. We can decompose every element of this subsequence into its $\mathbf{H}_2^{\perp}$ and $\mathbf{H}_2$ components, and choose a subsequence so that the first component converges. Proceeding in this way, and then using the Cantor diagonal trick of choosing the k-th term of the k-th selected subsequence, we have found a subsequence such that for any fixed j, the (now relabeled) subsequence, the $\mathbf{H}_j^{\perp}$ component of Tu_j converges. But the $\mathbf{H}_j$ component of Tu_j has norm less than $1/j$, and so the sequence converges by the triangle inequality.

2.4 Fourier's Fourier series.

We want to apply the theorem about compact self-adjoint operators that we proved in the preceding section to conclude that the functions e^{inx} form an orthonormal basis of the space $\mathcal{C}(\mathbf{T})$. In fact, a direct proof of this fact is elementary, using integration by parts. So we will pause to given this direct proof. Then we will go back and give a (more complicated) proof of the same fact using our theorem on compact operators. The reason for giving the more complicated proof is that it extends to far more general situations.

2.4.1 Proof by integration by parts.

We have let $\mathcal{C}(\mathbf{T})$ denote the space of continuous functions on the real line which are periodic with period 2π. We will let $\mathcal{C}^1(\mathbf{T})$ denote the space of periodic functions which have a continuous first derivative (necessarily periodic) and by $\mathcal{C}^2(\mathbf{T})$ the space of periodic functions with two continuous derivatives. If f and g both belong to $\mathcal{C}^1(\mathbf{T})$ then integration by parts gives

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f'\overline{g}dx = -\frac{1}{2\pi}\int_{-\pi}^{\pi} f\overline{g}'dx$$

since the boundary terms, which normally arise in the integration by parts formula, cancel, due to the periodicity of f and g. If we take $g = e^{inx}/(in), n \neq 0$ the integral on the right hand side of this equation is the Fourier coefficient:

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx.$$

We thus obtain

$$c_n = \frac{1}{in}\frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)e^{-inx}dx$$

so, for $n \neq 0$,

$$|c_n| \leq \frac{A}{n} \quad \text{where } A := \frac{1}{2\pi}\int_{-\pi}^{\pi} |f'(x)|dx$$

is a constant independent of n (but depending on f).

If $f \in \mathcal{C}^2(\mathbf{T})$ we can take $g(x) = -e^{inx}/n^2$ and integrate by parts twice. We conclude that (for $n \neq 0$)

$$|c_n| \leq \frac{B}{n^2} \quad \text{where} \quad B := \frac{1}{2\pi}\int_{-\pi}^{\pi} |f''(x)|^2$$

is again independent of n. But this proves that the Fourier series of f,

$$\sum c_n e^{inx}$$

converges uniformly and absolutely for and $f \in \mathcal{C}^2(\mathbf{T})$. The limit of this series is therefore some continuous periodic function. We must prove that this limit equals f. So we must prove that at each point f

$$\sum c_n e^{iny} \to f(y).$$

Replacing $f(x)$ by $f(x-y)$ it is enough to prove this formula for the case $y = 0$. So we must prove that for any $f \in \mathcal{C}^2(\mathbf{T})$ we have

$$\lim_{N,M\to\infty} \sum_{-N}^{M} c_n \to f(0).$$

Write $f(x) = (f(x) - f(0)) + f(0)$. The Fourier coefficients of any constant function c all vanish except for the c_0 term which equals c. So the above limit is trivially true when f is a constant. Hence, in proving the above formula, it is enough to prove it under the additional assumption that $f(0) = 0$, and we need to prove that in this case

$$\lim_{N,M\to\infty} (c_{-N} + c_{-N+1} + \cdots + c_M) \to 0.$$

The expression in parenthesis is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g_{N,M}(x)}dx$$

where

$$g_{N,M}(x) = e^{-iNx} + e^{-i(N-1)x} + \cdots + e^{iMx} = e^{-iNx}\left(1 + e^{ix} + \cdots + e^{i(M+N)x}\right) =$$

$$e^{-iNx}\frac{1 - e^{i(M+N+1)x}}{1 - e^{ix}} = \frac{e^{-iNx} - e^{i(M+1)x}}{1 - e^{ix}}, \quad x \neq 0$$

where we have used the formula for a geometric sum. By l'Hôpital's rule, this extends continuously to the value $M + N + 1$ for $x = 0$. Now $f(0) = 0$, and since f has two continuous derivatives, the function

$$h(x) := \frac{f(x)}{1 - e^{-ix}}$$

defined for $x \neq 0$ (or any multiple of 2π) extends, by l'Hôpital's rule, to a function defined at all values, and which is continuously differentiable and periodic. Hence the limit we are computing is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} h(x)e^{iNx}dx - \frac{1}{2\pi}\int_{-\pi}^{\pi} h(x)e^{-i(M+1)x}dx$$

and we know that each of these terms tends to zero.

We have thus proved that the Fourier series of any twice differentiable periodic function converges uniformly and absolutely to that function. If we consider the space $\mathcal{C}^2(\mathbf{T})$ with our usual scalar product

$$(f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\overline{g}dx$$

then the functions e^{inx} are dense in this space, since uniform convergence implies convergence in the $\| \ \|$ norm associated to $(\ , \ )$. So, on general principles, Bessel's inequality and Parseval's equation hold.

It is not true in general that the Fourier series of a continuous function converges uniformly to that function (or converges at all in the sense of uniform convergence). However it is true that we *do* have convergence in the L_2 norm, i.e. the Hilbert space $\| \ \|$ norm on $\mathcal{C}(\mathbf{T})$. To prove this, we need only prove that the exponential functions e^{inx} are dense, and since they are dense in $\mathcal{C}^2(\mathbf{T})$, it is enough to prove that $\mathcal{C}^2(\mathbf{T})$ is dense in $\mathcal{C}(\mathbf{T})$. For this, let ϕ be a function defined on the line with at least two continuous bounded derivatives with $\phi(0) = 1$ and of total integral equal to one and which vanishes rapidly at infinity. A favorite is the Gauss normal function

$$\phi(x) := \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

Equally well, we could take ϕ to be a function which actually vanishes outside of some neighborhood of the origin. Let

$$\phi_t(x) := \frac{1}{t}\phi\left(\frac{x}{t}\right).$$

As $t \to 0$ the function ϕ_t becomes more and more concentrated about the origin, but still has total integral one. Hence, for any bounded continuous function f, the function $\phi_t \star f$ defined by

$$(\phi_t \star f)(x) := \int_{-\infty}^{\infty} f(x-y)\phi_t(y)dy = \int_{-\infty}^{\infty} f(u)\phi_t(x-u)du.$$

satisfies $\phi_t \star f \to f$ uniformly on any finite interval. From the rightmost expression for $\phi_t \star f$ above we see that $\phi_t \star f$ has two continuous derivatives. From the first expression we see that $\phi_t \star f$ is periodic if f is. This proves that $\mathcal{C}^2(\mathbf{T})$ is dense in $\mathcal{C}(\mathbf{T})$. We have thus proved convergence in the L_2 norm.

2.4.2 Relation to the operator $\frac{d}{dx}$.

Each of the functions $e^{inx}$is an eigenvector of the operator

$$D = \frac{d}{dx}$$

in that

$$D\left(e^{inx}\right) = ine^{inx}.$$

So they are also eigenvalues of the operator D^2 with eigenvalues $-n^2$. Also, on the space of twice differentiable periodic functions the operator D^2 satisfies

$$(D^2 f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f''(x)\overline{g(x)}dx = \left. f'(x)\overline{g(x)}\right|_{-\pi}^{\pi} - \frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)\overline{g'(x)}dx$$

by integration by parts. Since f' and g are assumed to be periodic, the end point terms cancel, and integration by parts once more shows that

$$(D^2 f, g) = (f, D^2 g) = -(f', g').$$

But of course D and certainly D^2 is not defined on $\mathcal{C}(\mathbf{T})$ since some of the functions belonging to this space are not differentiable. Furthermore, the eigenvalues of D^2 are tending to infinity rather than to zero. So somehow the operator D^2 must be replaced with something like its inverse. In fact, we will work with the inverse of $D^2 - 1$, but first some preliminaries.

We will let $\mathcal{C}^2([-\pi, \pi])$ denote the functions defined on $[-\pi.\pi]$ and twice differentiable there, with continuous second derivatives up to the boundary. We denote by $\mathcal{C}([-\pi, \pi])$ the space of functions defined on $[-\pi, \pi]$ which are continuous up to the boundary. We can regard $\mathcal{C}(\mathbf{T})$ as the subspace of $\mathcal{C}([-\pi, \pi])$ consisting of those functions which satisfy the boundary conditions $f(\pi) = f(-\pi)$ (and then extended to the whole line by periodicity).

We regard $\mathcal{C}([-\pi, \pi])$ as a pre-Hilbert space with the same scalar product that we have been using:

$$(f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

If we can show that every element of $\mathcal{C}([-\pi,\pi])$ is a sum of its Fourier series (in the pre-Hilbert space sense) then the same will be true for $\mathcal{C}(\mathbf{T})$. So we will work with $\mathcal{C}([-\pi,\pi])$.

We can consider the operator D^2-1 as a linear map

$$D^2-1:\mathcal{C}^2([-\pi,\pi])\to\mathcal{C}([-\pi,\pi]).$$

This map is surjective, meaning that given any continuous function g we can find a twice differentiable function f satisfying the differential equation

$$f''-f=g.$$

In fact we can find a whole two dimensional family of solutions because we can add any solution of the homogeneous equation

$$h''-h=0$$

to f and still obtain a solution. We could write down an explicit solution for the equation $f''-f=g$, but we will not need to. It is enough for us to know that the solution exists, which follows from the general theory of ordinary differential equations.

The general solution of the homogeneous equation is given by

$$h(x)=ae^x+be^{-x}.$$

Let

$$M\subset\mathcal{C}^2([-\pi,\pi])$$

be the subspace consisting of those functions which satisfy the "periodic boundary conditions"

$$f(\pi)=f(-\pi),\quad f'(\pi)=f'(-\pi).$$

Given any f we can always find a solution of the homogeneous equation such that $f-h\in M$. Indeed, we need to choose the complex numbers a and b such that if h is as given above, then

$$h(\pi)-h(-\pi)=f(\pi)-f(-\pi),\ \text{and}\ \ h'(\pi)-h'(-\pi)=f'(\pi)-f'(-\pi).$$

Collecting coefficients and denoting the right hand side of these equations by c and d we get the linear equations

$$(e^\pi-e^{-\pi})(a-b)=c,(e^\pi-e^{-\pi})(a+b)=d$$

which has a unique solution.

So there exists a unique operator

$$T:\mathcal{C}([-\pi,\pi])\to M$$

with the property that

$$(D^2-I)\circ T=I.$$

We will prove that

$$T \text{ is self adjoint and compact.} \tag{2.20}$$

Once we will have proved this fact, then we know every element of M can be expanded in terms of a series consisting of eigenvectors of T with non-zero eigenvalues. But if

$$Tw = \lambda w$$

then

$$D^2 w = (D^2 - I)w + w = \frac{1}{\lambda}[(D^2 - I) \circ T]w + w = \left(\frac{1}{\lambda} + 1\right) w.$$

So w must be an eigenvector of D^2; it must satisfy

$$w'' = \mu w.$$

So if $\mu = 0$ then $w =$ a constant is a solution. If $\mu = r^2 > 0$ then w is a linear combination of e^{rx} and e^{-rx} and as we showed above, no non-zero such combination can belong to M. If $\mu = -r^2$ then the solution is a linear combination of e^{irx} and e^{-irx} and the above argument shows that r must be such that $e^{ir\pi} = e^{-ir\pi}$ so $r = n$ is an integer.

Thus (2.20) will show that the e^{inx} are a basis of M, and a little more work that we will do at the end will show that they are in fact also a basis of $\mathcal{C}([-\pi, \pi])$. But first let us work on (2.20).

It is easy to see that T is self adjoint. Indeed, let $f = Tu$ and $g = Tv$ so that f and g are in M and

$$(u, Tv) = ([D^2 - 1]f, g) = -(f', g') - (f, g) = (f, [D^2 - 1]g) = (Tu, v)$$

where we have used integration by parts and the boundary conditions defining M for the two middle equalities.

2.4.3 Gårding's inequality, special case.

We now turn to the compactness. We have already verified that for any $f \in M$ we have

$$([D^2 - 1]f, f) = -(f', f') - (f, f).$$

Taking absolute values we get

$$\|f'\|^2 + \|f\|^2 \leq |([D^2 - 1]f, f)|. \tag{2.21}$$

(We actually get equality here, the more general version of this that we will develop later will be an inequality.)

Let $u = [D^2 - 1]f$ and use the Cauchy-Schwartz inequality

$$|([D^2 - 1]f, f)| = |(u, f)| \leq \|u\|\|f\|$$

on the right hand side of (2.21) to conclude that

$$\|f\|^2 \leq \|u\|\|f\|$$

or

$$\|f\| \leq \|u\|.$$

Use (2.21) again to conclude that

$$\|f'\|^2 \leq \|u\|\|f\| \leq \|u\|^2$$

by the preceding inequality. We have $f = Tu$, and let us now suppose that u lies on the unit sphere i.e. that $\|u\| = 1$. Then we have proved that

$$\|f\| \leq 1, \quad \text{and} \quad \|f'\| \leq 1. \tag{2.22}$$

We wish to show that from any sequence of functions satisfying these two conditions we can extract a subsequence which converges. Here convergence means, of course, with respect to the norm given by

$$\|f\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx.$$

In fact, we will prove something stronger: that given any sequence of functions satisfying (2.22) we can find a subsequence which converges in the uniform norm

$$\|f\|_\infty := \max_{x\in[-\pi,\pi]} |f(x)|.$$

Notice that

$$\|f\| = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx\right)^{\frac{1}{2}} \leq \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} (\|f\|_\infty)^2 dx\right)^{\frac{1}{2}} = \|f\|_\infty$$

so convergence in the uniform norm implies convergence in the norm we have been using.

To prove our result, notice that for any $\pi \leq a < b \leq \pi$ we have

$$|f(b) - f(a)| = \left|\int_a^b f'(x)dx\right| \leq \int_a^b |f'(x)|dx = 2\pi(|f'|, \mathbf{1}_{[a,b]})$$

where $\mathbf{1}_{[a,b]}$ is the function which is one on $[a,b]$ and zero elsewhere. Apply Cauchy-Schwartz to conclude that

$$|(|f'|, \mathbf{1}_{[a,b]})| \leq \| \, |f'| \, \| \cdot \|\mathbf{1}_{[a,b]}\|.$$

But

$$\|\mathbf{1}_{[a,b]}\|^2 = \frac{1}{2\pi}|b-a|$$

and

$$\| \, |f'| \| = \|f'\| \leq 1.$$

We conclude that

$$|f(b) - f(a)| \leq (2\pi)^{\frac{1}{2}} |b - a|^{\frac{1}{2}}. \tag{2.23}$$

In this inequality, let us take b to be a point where $|f|$ takes on its maximum value, so that $|f(b)| = \|f\|_\infty$. Let a be a point where $|f|$ takes on its minimum value. (If necessary interchange the role of a and b to arrange that $a < b$ or observe that the condition $a < b$ was not needed in the above proof.) Then (2.23) implies that

$$\|f\|_\infty - \min |f| \leq (2\pi)^{\frac{1}{2}} |b - a|^{\frac{1}{2}}.$$

But

$$1 \geq \|f\| = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f|^2(x) dx \right)^{\frac{1}{2}} \geq \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (\min |f|)^2 dx \right)^{\frac{1}{2}} = \min |f|$$

and $|b - a| \leq 2\pi$ so

$$\|f\|_\infty \leq 1 + 2\pi.$$

Thus the values of all the $f \in T[S]$ are all uniformly bounded - (they take values in a circle of radius $1 + 2\pi$) and they are equicontinuous in that (2.23) holds. This is enough to guarantee that out of every sequence of such f we can choose a uniformly convergent subsequence.

(We recall how the proof of this goes: Since all the values of all the f are bounded, at any point we can choose a subsequence so that the values of the f at that point converge, and, by passing to a succession of subsequences (and passing to a diagonal), we can arrange that this holds at any countable set of points. In particular, we may choose say the rational points in $[-\pi, \pi]$. Suppose that f_n is this subsequence. We claim that (2.23) then implies that the f_n form a Cauchy sequence in the uniform norm and hence converge in the uniform norm to some continuous function. Indeed, for any ϵ choose δ such that

$$(2\pi)^{\frac{1}{2}} \delta^{\frac{1}{2}} < \frac{1}{3}\epsilon,$$

choose a finite number of rational points which are within δ distance of any point of $[-\pi, \pi]$ and choose N sufficiently large that $|f_i - f_j| < \frac{1}{3}\epsilon$ at each of these points, r. when i and j are $\geq N$. Then at any $x \in [-\pi, \pi]$

$$|f_i(x) - f_j(x)| \leq |f_i(x) - f_i(r)| + |f_j(x) - f_j(r)| + |f_i(r) - f_j(r)| \leq \epsilon$$

since we can choose r such that that the first two and hence all of the three terms is $\leq \frac{1}{3}\epsilon$.)

2.5 The Heisenberg uncertainty principle.

In this section we show how the arguments leading to the Cauchy-Schwartz inequality give one of the most important discoveries of twentieth century physics, the Heisenberg uncertainty principle.

Let V be a pre-Hilbert space, and S denote the unit sphere in V. If ϕ and ψ are two unit vectors (i.e. elements of S) their scalar product (ϕ, ψ) is a complex number with $0 \leq |(\phi, \psi)|^2 \leq 1$. In quantum mechanics, this number is taken as a probability. Although in the "real world" V is usually infinite dimensional, we will warm up by considering the case where V is finite dimensional.

Given a $\phi \in S$ and an orthonormal basis $\phi_1, \ldots, \phi_n$ of V, we have

$$1 = \|\phi\|^2 = |(\phi, \phi_1)|^2 + \cdots + |(\phi, \phi_n|^2.$$

The says that the various probabilities $|(\phi, \phi_i)|^2$ add up to one. We recall some language from elementary probability theory: Suppose we have an experiment resulting in a finite number of measured numerical outcomes λ_i, each with probability p_i of occurring. Then the mean $\langle\lambda\rangle$ is defined by

$$\langle\lambda\rangle := \lambda_1 p_1 + \cdots + \lambda_n p_n$$

and its variance $(\Delta\lambda)^2$

$$(\Delta\lambda)^2 := (\lambda_1 - \langle\lambda\rangle)^2 p_1 + \cdots + (\lambda_n - \langle\lambda\rangle)^2 p_n$$

and an immediate computation shows that

$$(\Delta\lambda)^2 = \langle\lambda^2\rangle - \langle\lambda\rangle^2.$$

The square root $\Delta\lambda$ of the variance is called the standard deviation. The variance (or the standard deviation) measures the "spread" of the possible values of λ. To understand its meaning we have Chebychev's inequality which estimates the probability that λ_k can deviate from $\langle\lambda\rangle$ by as much as $r\Delta\lambda$ for any positive number r. Chebychev's inequality says that this probability is $\leq 1/r^2$. In symbols

$$\text{Prob } (|\lambda_k - \langle\lambda\rangle| \geq r\Delta\lambda) \leq \frac{1}{r^2}.$$

Indeed, the probability on the left is the sum of all the p_k such that $|\lambda_k - \langle\lambda\rangle| \geq r\Delta$. Denoting this sum by $\sum_r$ we have

$$\sum\nolimits_r p_k \leq \sum\nolimits_r p_k \frac{(\lambda - \langle\lambda\rangle)^2}{r^2(\Delta\lambda)^2} \leq$$

$$\leq \sum_{\text{all } k} p_k \frac{(\lambda - \langle\lambda\rangle)^2}{r^2(\Delta\lambda)^2} = \frac{1}{r^2(\Delta\lambda)^2} \sum_{\text{all } k} (\lambda - \langle\lambda\rangle)^2 p_k = \frac{1}{r^2}.$$

Replacing λ_i by $\lambda_i + c$ does not change the variance.

Now suppose that A is a self-adjoint operator on V, that the λ_i are the eigenvalues of A with eigenvectors ϕ_i constituting an orthonormal basis, and that the $p_i = |(\phi, \phi_i|^2$ as above.

1. Show that $\langle\lambda\rangle = (A\phi, \phi)$ and that $(\Delta\lambda)^2 = (A^2\phi, \phi) - (A\phi, \phi)^2$.

We will write the expression $(A\phi, \phi)$ as $\langle A\rangle_\phi$, In quantum mechanics a unit vector is called a **state** and a self-adjoint operator is called an **observable** and the expression $\langle A\rangle_\phi$ is called the **expectation** of the observable A in the state ϕ. Similarly we denote $\left((A^2\phi, \phi) - (A\phi, \phi)^2\right)^{1/2}$ by $\Delta_\phi A$. It is called the **uncertainty** of the observable A in the state ϕ. Notice that

$$(\Delta_\phi A)^2 = \langle (A - \langle A\rangle I)^2\rangle_\phi$$

where I denotes the identity operator. Indeed

$$((A - (A\phi, \phi)I)^2 = A^2 - 2(A\phi, \phi)A + (A\phi, \phi)^2 I$$

so

$$\langle (A - \langle A\rangle_\phi)^2\rangle_\phi = (A^2\phi, \phi) - 2\langle A\rangle_\phi^2 + \langle A\rangle_\phi^2 = \langle A^2\rangle_\phi - \langle A\rangle_\phi^2.$$

When the state ϕ is fixed in the course of discussion, we will drop the subscript ϕ and write $\langle A\rangle$ and ΔA instead of $\langle A\rangle_\phi$ and $\Delta_\phi A$. For example, we would write the previous result as

$$\Delta A = \langle (A - \langle A\rangle I)^2\rangle.$$

If A and B are operators we let $[A, B]$ denote the commutator:

$$[A, B] := AB - BA.$$

Notice that $[A, B] = -[B, A]$ and $[I, B] = 0$ for any B. So if A and B are self adjoint, so is $i[A, B]$ and replacing A by $A - \langle A\rangle I$ and B by $B - \langle B\rangle I$ does not change ΔA, ΔB or $i[A, B]$.

The **uncertainty principle** says that for any two observables A and B we have

$$(\Delta A)(\Delta B) \geq \frac{1}{2}|\langle i[A, B]\rangle| q.$$

Proof. Set $A_1 := A - \langle A\rangle, \quad B_1 := B - \langle B\rangle$ so that

$$[A_1, B_1] = [A, B].$$

Let

$$\psi := A_1\phi + ixB_1\phi.$$

Then

$$(\psi, \psi) = (\Delta A)^2 - x\langle i[A, B]\rangle + (\Delta B)^2.$$

Since $(\psi, \psi) \geq 0$ for all x this implies that ($b^2 \leq 4ac$) that

$$\langle i[A, B]\rangle^2 \leq 4(\Delta A)^2(\Delta B)^2,$$

and taking square roots gives the result.

The purpose of the next few sections is to provide a vast generalization of the results we obtained for the operator D^2. We will prove the corresponding results for any "elliptic" differential operator (definitions below).

I plan to study differential operators acting on vector bundles over manifolds. But it requires some effort to set things up, and I want to get to the key analytic ideas which are essentially repeated applications of integration by parts. So I will start with elliptic operators L acting on functions on the torus $\mathbf{T} = \mathbf{T}^n$, where there are no boundary terms when we integrate by parts. Then an immediate extension gives the result for elliptic operators on functions on manifolds, and also for boundary value problems such as the Dirichlet problem.

The treatment here rather slavishly follows the treatment by Bers and Schechter in *Partial Differential Equations* by Bers, John and Schechter AMS (1964).

2.6 The Sobolev Spaces.

Recall that $\mathbf{T}$ now stands for the n-dimensional torus. Let $\mathbf{P} = \mathbf{P}(\mathbf{T})$ denote the space of trigonometric polynomials. These are functions on the torus of the form

$$u(x) = \sum a_\ell e^{i\ell\cdot x}$$

where

$$\ell = (\ell_1, \dots, \ell_n)$$

is an n-tuplet of integers and the sum is finite. For each integer t (positive, zero or negative) we introduce the scalar product

$$(u, v)_t := \sum_\ell (1 + \ell\cdot\ell)^t a_\ell \bar{b}_\ell. \tag{2.24}$$

For $t = 0$ this is the scalar product

$$(u, v)_0 = \frac{1}{(2\pi)^n}\int_{\mathbf{T}} u(x)\overline{v(x)}dx.$$

This differs by a factor of $(2\pi)^{-n}$ from the scalar product that is used by Bers and Schecter. We will denote the norm corresponding to the scalar product $(\ ,\ )_s$ by $\|\ \|_s$.

If

$$\Delta := -\left(\frac{\partial^2}{\partial(x^1)^2} + \cdots + \frac{\partial^2}{\partial(x^n)^2}\right)$$

the operator $(1 + \Delta)$ satisfies

$$(1 + \Delta)u = \sum (1 + \ell\cdot\ell)a_\ell e^{i\ell\cdot x}$$

and so

$$((1 + \Delta)^t u, v)_s = (u, (1 + \Delta)^t v)_s = (u, v)_{s+t}$$

and

$$\|(1 + \Delta)^t u\|_s = \|u\|_{s+2t}. \tag{2.25}$$

We then get the "generalized Cauchy-Schwartz inequality"

$$|(u,v)_s| \leq \|u\|_{s+t}\|v\|_{s-t} \tag{2.26}$$

for any t, as a consequence of the usual Cauchy-Schwartz inequality. Indeed,

$$\begin{aligned}
\sum_\ell (1+\ell\cdot\ell)^s a_\ell \overline{b}_\ell &= \sum_\ell (1+\ell\cdot\ell)^{\frac{s+t}{2}} a_\ell (1+\ell\cdot\ell)^{\frac{s-t}{2}} \overline{b}_\ell \\
&= ((1+\Delta)^{\frac{s+t}{2}} u, (1+\Delta)^{\frac{s-t}{2}} v)_0 \\
&\leq \|(1+\Delta)^{\frac{s+t}{2}} u\|_0 \|(1+\Delta)^{\frac{s-t}{2}} v\|_0 \\
&= \|u\|_{s+t}\|v\|_{s-t}.
\end{aligned}$$

The generalized Cauchy-Schwartz inequality reduces to the usual Cauchy-Schwartz inequality when $t = 0$.

Clearly we have

$$\|u\|_s \leq \|u\|_t \quad \text{if } s \leq t.$$

If D^p denotes a partial derivative,

$$D^p = \frac{\partial^{|p|}}{\partial (x^1)^{p_1} \cdots \partial (x^n)^{p_m}}$$

then

$$D^p u = \sum (i\ell)^p a_\ell e^{i\ell\cdot x}.$$

In these equations we are using the following notations:

- If $p = (p_1, \ldots, p_n)$ is a vector with non-negative integer entries we set

$$|p| := p_1 + \cdots + p_n.$$

- If $\xi = (\xi_1, \ldots, \xi_n)$ is a (row) vector we set

$$\xi^p := \xi_1^{p_1} \cdot \xi_2^{p_2} \cdots \xi_n^{p_n}$$

It is then clear that

$$\|D^p u\|_t \leq \|u\|_{t+|p|} \tag{2.27}$$

and similarly

$$\|u\|_t \leq (\text{constant depending on } t) \sum_{|p|\leq t} \|D^p u\|_0 \quad \text{if } t \geq 0. \tag{2.28}$$

In particular,

Proposition 2.6.1 *The norms*

$$u \mapsto \|u\|_t$$

$t \geq 0$ *and*

$$u \mapsto \sum_{|p| \leq t} \|D^p u\|_0$$

are equivalent.

We let $\mathbf{H}_t$ denote the completion of the space $\mathbf{P}$ with respect to the norm $\| \ \|_t$. Each $\mathbf{H}_t$ is a Hilbert space, and we have natural embeddings

$$\mathbf{H}_t \hookrightarrow \mathbf{H}_s \ \text{ if } s < t.$$

Equation (2.25) says that

$$(1 + \Delta)^t : \mathbf{H}_{s+2t} \to \mathbf{H}_s$$

and is an isometry.

From the generalized Schwartz inequality we also have a natural pairing of $\mathbf{H}_t$ with $\mathbf{H}_{-t}$ given by the extension of $(\ ,\ )_0$, so

$$|(u, v)_0| \leq \|u\|_t \|v\|_{-t}. \tag{2.29}$$

In fact, this pairing allows us to identify $\mathbf{H}_{-t}$ with the space of continuous linear functions on $\mathbf{H}_t$. Indeed, if ϕ is a continuous linear function on $\mathbf{H}_t$ the Riesz representation theorem tells us that there is a $w \in \mathbf{H}_t$ such that $\phi(u) = (u, w)_t$. Set

$$v := (1 + \Delta)^t w.$$

Then

$$v \in \mathbf{H}_{-t}$$

and

$$(u, v)_0 = (u, (1 + \Delta)^t w)_0 = (u, w)_t = \phi(u).$$

We record this fact as

$$\mathbf{H}_{-t} = (\mathbf{H}_t)^* . \tag{2.30}$$

As an illustration of (2.30), observe that the series

$$\sum_\ell (1 + \ell \cdot \ell)^s$$

converges for

$$s < -\frac{n}{2}.$$

This means that if define v by taking

$$b_\ell \equiv 1$$

then $v \in \mathbf{H}_s$ for $s < -\frac{n}{2}$. If u is given by $u(x) = \sum_\ell a_\ell e^{i\ell \cdot x}$ is any trigonometric polynomial, then

$$(u, v)_0 = \sum a_\ell = u(0).$$

So the natural pairing (2.29) allows us to extend the linear function sending $u \mapsto u(0)$ to all of $\mathbf{H}_t$ if $t > \frac{n}{2}$. We can now give v its "true name": it is the Dirac "delta function" δ (on the torus) where

$$(u, \delta)_0 = u(0).$$

So $\delta \in H_{-t}$ for $t > \frac{n}{2}$, and the preceding equation is usually written symbolically as

$$\frac{1}{(2\pi)^n} \int_{\mathbf{T}} u(x)\delta(x)dx = u(0);$$

but the true mathematical interpretation is as given above.

We set

$$\mathbf{H}_\infty := \bigcap \mathbf{H}_t, \quad \mathbf{H}_{-\infty} := \bigcup \mathbf{H}_t.$$

The space $\mathbf{H}_0$ is just $L_2(\mathbf{T})$, and we can think of the space $\mathbf{H}_t, \quad t > 0$ as consisting of those functions having "generalized L_2 derivatives up to order t". Certainly a function of class C^t belongs to $\mathbf{H}_t$. With a loss of degree of differentiability the converse is true:

Lemma 2.6.1 [Sobolev.] *If $u \in \mathbf{H}_t$ and*

$$t \geq \left[\frac{n}{2}\right] + k + 1$$

then $u \in C^k(\mathbf{T})$ and

$$\sup_{x \in \mathbf{T}} |D^p u(x)| \leq \text{const.} \|u\|_t \quad \textit{for } |p| \leq k. \tag{2.31}$$

By applying the lemma to $D^p u$ it is enough to prove the lemma for $k = 0$. So we assume that $u \in \mathbf{H}_t$ with $t \geq [n/2] + 1$. Then

$$(\sum |a_\ell|)^2 \leq \left(\sum (1 + \ell \cdot \ell)^t |a_\ell|^2\right) \sum (1 + \ell \cdot \ell)^{-t} < \infty,$$

since the series $\sum (1 + \ell \cdot \ell)^{-t}$ converges for $t \geq [n/2] + 1$. So for this range of t, the Fourier series for u converges absolutely and uniformly. The right hand side of the above inequality gives the desired bound. QED

A **distribution** on $\mathbf{T}^n$ is a linear function T on $C^\infty(\mathbf{T}^n)$ with the continuity condition that

$$\langle T, \phi_k \rangle \to 0$$

whenever

$$D^p \phi_k \to 0$$

uniformly for each fixed p. If $u \in \mathbf{H}_{-t}$ we may define

$$\langle u, \phi \rangle := (\phi, \overline{u})_0$$

and since $C^\infty(\mathbf{T})$ is dense in $\mathbf{H}_t$ we may conclude

Lemma 2.6.2 $\mathbf{H}_{-t}$ *is the space of those distributions* T *which are continuous in the* $\| \ \|_t$ *norm, i.e. which satisfy*

$$\|\phi_k\|_t \to 0 \quad \Rightarrow \ \langle T, \phi_k \rangle \to 0.$$

We then obtain

Theorem 2.6.1 [Laurent Schwartz.] $\mathbf{H}_{-\infty}$ *is the space of all distributions. In other words, any distribution belongs to* $\mathbf{H}_{-t}$ *for some* t.

Proof. Suppose that T is a distribution that does not belong to any $\mathbf{H}_{-t}$. This means that for any $k > 0$ we can find a C^∞ function ϕ_k with

$$\|\phi_k\|_k < \frac{1}{k}$$

and

$$|\langle T, \phi_k \rangle| \geq 1.$$

But by Lemma 2.6.1 we know that $\|\phi_k\|_k < \frac{1}{k}$ implies that $D^p \phi_k \to 0$ uniformly for any fixed p contradicting the continuity property of T. QED

Suppose that ϕ is a C^∞ function on $\mathbf{T}$. Multiplication by ϕ is clearly a bounded operator on $\mathbf{H}_0 = L_2(\mathbf{T})$, and so it is also a bounded operator on $\mathbf{H}_t, \ \ t > 0$ since we can expand $D^p(\phi u)$ by applications of Leibnitz's rule.

For $t = -s < 0$ we know by the generalized Cauchy Schwartz inequality that

$$\|\phi u\|_t = \sup |(v, \phi u)_0| / \|v\|_s = \sup |(u, \overline{\phi} v)| / \|v\|_s \leq \|u\|_t \|\overline{\phi} v\|_s / \|v\|_s.$$

So in all cases we have

$$\|\phi u\|_t \leq (\text{const. depending on } \phi \text{ and } t) \|u\|_t. \tag{2.32}$$

Let

$$L = \sum_{|p| \leq m} \alpha_p(x) D^p$$

be a differential operator of degree m with C^∞ coefficients. Then it follows from the above that

$$\|Lu\|_{t-m} \leq \text{constant} \|u\|_t \tag{2.33}$$

where the constant depends on L and t.

Lemma 2.6.3 [Rellich's lemma.] *If* $s < t$ *the embedding* $\mathbf{H_t} \hookrightarrow \mathbf{H}_s$ *is compact.*

Proof. We must show that the image of the unit ball B of $\mathbf{H}_t$ in $\mathbf{H}_s$ can be covered by finitely many balls of radius ϵ. Choose N so large that

$$(1+\ell\cdot\ell)^{(s-t)/2} < \frac{\epsilon}{2}$$

when $\ell\cdot\ell > N$. Let Z_t be the subspace of $\mathbf{H}_t$ consisting of all u such that $a_\ell = 0$ when $\ell\cdot\ell \leq N$. This is a space of finite codimension, and hence the unit ball of $Z_t^\perp \subset \mathbf{H}_t$ can be covered by finitely many balls of radius $\frac{\epsilon}{2}$. The space $Z_t^\perp$ consists of all u such that $a_\ell = 0$ when $\ell\cdot\ell > N$. The image of $Z_t^\perp$ in $\mathbf{H_s}$ is the orthogonal complement of the image of Z_t. On the other hand, for $u \in B \cap Z$ we have

$$\|u\|_s^2 \leq (1+N)^{s-t}\|u\|_t^2 \leq \left(\frac{\epsilon}{2}\right)^2.$$

So the image of $B\cap Z$ is contained in a ball of radius $\frac{\epsilon}{2}$. Every element of the image of B can be written as a(n orthogonal) sum of an element in the image of $B\cap Z_t^\perp$ and an element of $B\cap Z_t$ and so the image of B is covered by finitely many balls of radius ϵ. QED

2.7 Gårding's inequality.

Let $x,\ a$, and b be positive numbers. Then

$$x^a + x^{-b} \geq 1$$

because if $x \geq 1$ the first summand is ≥ 1 and if $x \leq 1$ the second summand is ≥ 1. Setting $x = \epsilon^{1/a}A$ gives

$$1 \leq \epsilon A^a + \epsilon^{-b/a}A^{-b}$$

if ϵ and A are positive. Suppose that $t_1 > s > t_2$ and we set $a = t_1 - s,\ b = s - t_2$ and $A = 1+\ell\cdot\ell$. Then we get

$$(1+\ell\cdot\ell)^s \leq \epsilon(1+\ell\cdot\ell)^{t_1} + \epsilon^{-(s-t_2)/(t_1-s)}(1+\ell\cdot\ell)^{t_2}$$

and therefore

$$\|u\|_s \leq \epsilon\|u\|_{t_1} + \epsilon^{-(s-t_2)/(t_1-s)}\|u\|_{t_2} \quad \text{if } t_1 > s > t_2, \quad \epsilon > 0 \tag{2.34}$$

for all $u \in \mathbf{H}_{t_1}$. This elementary inequality will be the key to several arguments in this section where we will combine (2.34) with integration by parts.

A differential operator $L = \sum_{|p|\leq m}\alpha_p(x)D^p$ with real coefficients and m even is called **elliptic** if there is a constant $c > 0$ such that

$$(-1)^{m/2}\sum_{|p|=m} a_p(x)\xi^p \geq c(\xi\cdot\xi)^{m/2}. \tag{2.35}$$

In this inequality, the vector ξ is a "dummy variable". (Its true invariant significance is that it is a covector, i.e. an element of the cotangent space at x.) The

expression on the left of this inequality is called the **symbol** of the operator L. It is a homogeneous polynomial of degree m in the variable ξ whose coefficients are functions of x. The symbol of L is sometimes written as $\sigma(L)$ or $\sigma(L)(x,\xi)$. Another way of expressing condition (2.35) is to say that there is a positive constant c such that

$$\sigma(L)(x,\xi) \geq c \quad \text{for all } x \quad \text{and } \xi \text{ such that } \quad \xi \cdot \xi = 1.$$

We will assume until further notice that the operator L is elliptic and that m is a positive even integer.

Theorem 2.7.1 [Gårding's inequality.] *For every $u \in C^\infty(\mathbf{T})$ we have*

$$(u, Lu)_0 \geq c_1 \|u\|^2_{m/2} - c_2 \|u\|^2_0 \tag{2.36}$$

where c_1 and c_2 are constants depending on L.

Remark. If $u \in \mathbf{H}_{m/2}$, then both sides of the inequality make sense, and we can approximate u in the $\| \ \|_{m/2}$ norm by C^∞ functions. So once we prove the theorem, we conclude that it is also true for all elements of $\mathbf{H}_{m/2}$.

We will prove the theorem in stages:

1. When L is constant coefficient and homogeneous.
2. When L is homogeneous and approximately constant.
3. When the L can have lower order terms but the homogeneous part of L is approximately constant.
4. The general case.

Stage 1. $L = \sum_{|p|=m} \alpha_p D^p$ where the α_p are constants. Then

$$\begin{aligned}
(u, Lu)_0 &= \left(\sum a_\ell e^{i\ell\cdot x}, \sum_\ell \left(\sum_{|p|=m} \alpha_p (i\ell)^p \right) a_\ell e^{i\ell\cdot x} \right)_0 \\
&\geq c \sum_\ell (\ell \cdot \ell)^{m/2} |a_\ell|^2 \quad \text{by (2.35)} \\
&= c \sum [1 + (\ell \cdot \ell)^{m/2}] |a_\ell|^2 - c\|u\|^2_0 \\
&\geq cC\|u\|^2_{m/2} - c\|u\|_0
\end{aligned}$$

where

$$C = \sup_{r \geq 0} \frac{1 + r^{m/2}}{(1+r)^{m/2}}.$$

This takes care of stage 1.

Stage 2. $L = L_0 + L_1$ where L_0 is as in stage 1 and $L_1 = \sum_{|p|=m} \beta_p(x) D^p$ and

$$\max_{p,x} |\beta_p(x)| < \eta,$$

where η sufficiently small. (How small will be determined very soon in the course of the discussion.) We have

$$(u, L_0 u)_0 \geq c' \|u\|^2_{m/2} - c\|u\|^2_0$$

from stage 1.

We integrate $(u, L_1 u)_0$ by parts $m/2$ times. There are no boundary terms since we are on the torus. In integrating by parts some of the derivatives will hit the coefficients. Let us collect all the these terms as I_2. The other terms we collect as I_1, so

$$I_1 = \sum \int b_{p'+p''} D^{p'} u \overline{D^{p''} u} dx$$

where $|p'| = |p''| = m/2$ and $b_r = \pm \beta_r$. We can estimate this sum by

$$|I_1| \leq \eta \cdot \text{const.} \|u\|^2_{m/2}$$

and so will require that $\eta \cdot (\text{const.}) < c'$.

The remaining terms give a sum of the form

$$I_2 = \sum \int b_{p'q} D^{p'} u \overline{D^q u} dx$$

where $p' \leq m/2, q' < m/2$ so we have

$$|I_2| \leq \text{const.} \|u\|_{\frac{m}{2}} \|u\|_{\frac{m}{2}-1}.$$

Now let us take

$$s = \frac{m}{2} - 1, \ t_1 = \frac{m}{2}, \ t_2 = 0$$

in (2.34) which yields, for any $\epsilon > 0$,

$$\|u\|_{\frac{m}{2}-1} \leq \epsilon \|u\|_{\frac{m}{2}} + \epsilon^{-m/2} \|u\|_0.$$

Substituting this into the above estimate for I_2 gives

$$|I_2| \leq \epsilon \cdot \text{const.} \|u\|^2_{m/2} + \epsilon^{-m/2} \text{const.} \|u\|_{m/2} \|u\|_0.$$

For any positive numbers a, b and ζ the inequality $(\zeta a - \zeta^{-1} b)^2 \geq 0$ implies that $2ab \leq \zeta^2 a^2 + \zeta^{-2} b^2$. Taking $\zeta^2 = \epsilon^{\frac{m}{2}+1}$ we can replace the second term on the right in the preceding estimate for $|I_2|$ by

$$\epsilon^{-m-1} \cdot \text{const.} \|u\|^2_0$$

at the cost of enlarging the constant in front of $\|u\|^2_{\frac{m}{2}}$. We have thus established that

$$|I_1| \leq \eta \cdot (\text{const.})_1 \|u\|^2_{m/2}$$

where the constant depends only on m, and

$$|I_2| \leq \epsilon(\text{const.})_2 \|u\|^2_{m/2} + \epsilon^{-m-1}\text{const.}\|u\|^2_0$$

where the constants depend on L_1 but ϵ is at our disposal. So if $\eta(\text{const.})_1 < c'$ and we then choose ϵ so that $\epsilon(\text{const.})_2 < c' - \eta \cdot (\text{const.})_1$ we obtain Gårding's inequality for this case.

Stage 3. $L = L_0 + L_1 + L_2$ where L_0 and L_1 are as in stage 2, and L_2 is a lower order operator. Here we integrate by parts and argue as in stage 2.

Stage 4, the general case. Choose an open covering of T such that the variation of each of the highest order coefficients in each open set is less than the η of stage 1. (Recall that this choice of η depended only on m and the c that entered into the definition of ellipticity.) Thus, if v is a smooth function supported in one of the sets of our cover, the action of L on v is the same as the action of an operator as in case 3) on v, and so we may apply Gårding's inequality. Choose a finite subcover and a partition of unity $\{\phi_i\}$ subordinate to this cover. Write $\phi_i = \psi_i^2$ (where we choose the ϕ so that the ψ are smooth). So $\sum \psi_i^2 \equiv 1$. Now

$$(\psi_i u, L(\psi_i u))_0 \geq c''\|\psi_i u\|^2_{m/2} - \text{const.}\|\psi_i u\|^2_0$$

where c'' is a positive constant depending only on c, η, and on the lower order terms in L. We have

$$(u, Lu)_0 = \int (\sum \psi_i^2 u)\overline{Lu}dx = \sum (\psi_i u, L\psi_i u)_0 + R$$

where R is an expression involving derivatives of the ψ_i and hence lower order derivatives of u. These can be estimated as in case 2) above, and so we get

$$(u, Lu)_0 \geq c''' \sum \|\psi_i u\|^2_{m/2} - \text{const.}\|u\|^2_0 \tag{2.37}$$

since $\|\psi_i u\|_0 \leq \|u\|_0$. Now $\|u\|_{m/2}$ is equivalent, as a norm, to $\sum_{p \leq m/2} \|D^p u\|_0$ as we verified in the preceding section. Also

$$\sum \|D^p(\psi_i u)\|_0 = \sum \|\psi_i D^p u\|_0 + R'$$

where R' involves terms differentiating the ψ and so lower order derivatives of u. Hence

$$\sum \|\psi_i u\|^2_{m/2} \geq \text{pos. const.}\|u\|^2_{m/2} - \text{const.}\|u\|^2_0$$

by the integration by parts argument again. Hence by (2.37)

$$(u, Lu)_0 \geq c''' \sum \|\psi_i u\|^2_{m/2} - \text{const.}\|u\|^2_0$$

$$\geq \text{pos. const.}\|u\|^2_{m/2} - \text{const.}\|u\|^2_0$$

which is Gårding's inequality. QED

For the time being we will continue to study the case of the torus. But a look ahead is in order. In this last step of the argument, where we applied the partition of unity argument, we have really freed ourselves of the restriction of being on the torus. Once we make the appropriate definitions, we will then get Gårding's inequality for elliptic operators on manifolds. Furthermore, the consequences we are about to draw from Gårding's inequality will be equally valid in the more general setting.

2.8 Consequences of Gårding's inequality.

Proposition 2.8.1 *For every integer t there is a constant $c(t) = c(t, L)$ and a positive number $\Lambda = \Lambda(t, L)$ such that*

$$\|u\|_t \leq c(t)\|Lu + \lambda u\|_{t-m} \tag{2.38}$$

when

$$\lambda > \Lambda$$

for all smooth u, and hence for all $u \in \mathbf{H}_t$.

Proof. Let s be some non-negative integer. We will first prove (2.38) for $t = s + \frac{m}{2}$. We have

$$\|u\|_t\|Lu + \lambda u\|_{t-m} = \|u\|_t\|Lu + \lambda u\|_{s-\frac{m}{2}}$$

$$= \|u\|_t\|(1+\Delta)^s Lu + \lambda(1+\Delta)^s u\|_{-s-\frac{m}{2}}$$

$$\geq (u, (1+\Delta)^s Lu + \lambda(1+\Delta)^s u)_0$$

by the generalized Cauchy - Schwartz inequality (2.26).

The operator $(1+\Delta)^s L$ is elliptic of order $m + 2s$ so (2.25) and Gårding's inequality gives

$$(u, (1+\Delta)^s Lu + \lambda(1+\Delta)^s u)_0 \geq c_1\|u\|^2_{s+\frac{m}{2}} - c_2\|u\|^2_0 + \lambda\|u\|^2_s.$$

Since $\|u\|_s \geq \|u\|_0$ we can combine the two previous inequalities to get

$$\|u\|_t\|Lu + \lambda u\|_{t-m} \geq c_1\|u\|^2_t + (\lambda - c_2)\|u\|^2_0.$$

If $\lambda > c_2$ we can drop the second term and divide by $\|u\|_t$ to obtain (2.38).

We now prove the proposition for the case $t = \frac{m}{2} - s$ by the same sort of argument: We have

$$\|u\|_t\|Lu + \lambda u\|_{-s-\frac{m}{2}} = \|(1+\Delta)^{-s}u\|_{s+\frac{m}{2}}\|Lu + \lambda u\|_{-s-\frac{m}{2}}$$

$$\geq ((1+\Delta)^{-s}u, L(1+\Delta)^s(1+\Delta)^{-s}u + \lambda u)_0.$$

Now use the fact that $L(1+\Delta)^s$ is elliptic and Gårding's inequality to continue the above inequalities as

$$\geq c_1\|(1+\Delta)^{-s}u\|^2_{s+\frac{m}{2}} - c_2\|(1+\Delta)^{-s}u\|^2_0 + \lambda\|u\|^2_{-s}$$

$$= c_1\|u\|^2_t - c_2\|u\|^2_{-2s} + \lambda\|u\|^2_{-s} \geq c_1\|u\|^2_t$$

if $\lambda > c_2$. Again we may then divide by $\|u\|_t$ to get the result. QED

The operator $L + \lambda I$ is a bounded operator from $\mathbf{H}_t$ to $\mathbf{H}_{t-m}$ (for any t). Suppose we fix t and choose λ so large that (2.38) holds. Then (2.38) says that $(L+\lambda I)$ is invertible on its image, and bounded there with a bound independent of $\lambda > \Lambda$, and this image is a closed subspace of $\mathbf{H}_{t-m}$.

Let us show that this image is all of $\mathbf{H}_{t-m}$ for λ large enough. Suppose not, which means that there is some $w \in \mathbf{H}_{t-m}$ with

$$(w, Lu + \lambda u)_{t-m} = 0$$

for all $u \in \mathbf{H}_t$. We can write this last equation as

$$((1+\Delta)^{t-m}w, Lu + \lambda u)_0 = 0.$$

Integration by parts gives the adjoint differential operator L^* characterized by

$$(\phi, L\psi)_0 = (L^*\phi, \psi)_0$$

for all smooth functions ϕ and ψ, and by passing to the limit this holds for all elements of $\mathbf{H}_r$ for $r \geq m$. The operator L^* has the same leading term as L and hence is elliptic. So let us choose λ sufficiently large that (2.38) holds for L^* as well as for L. Now

$$0 = \big((1+\Delta)^{t-m}w, Lu + \lambda u\big)_0 = \big(L^*(1+\Delta)^{t-m}w + \lambda(1+\Delta)^{t-m}w, u\big)_0$$

for all $u \in \mathbf{H}_t$ which is dense in $\mathbf{H}_0$ so

$$L^*(1+\Delta)^{t-m}w + \lambda(1+\Delta)^{t-m}w = 0$$

and hence (by (2.38)) $(1+\Delta)^{t-m}w = 0$ so $w = 0$. We have proved

Proposition 2.8.2 *For every t and for λ large enough (depending on t) the operator $L + \lambda I$ maps $\mathbf{H}_t$ bijectively onto $\mathbf{H}_{t-m}$ and $(L+\lambda I)^{-1}$ is bounded independently of λ.*

As an immediate application we get the important

Theorem 2.8.1 *If u is a distribution and $Lu \in \mathbf{H}_s$ then $u \in \mathbf{H}_{s+m}$.*

Proof. Write $f = Lu$. By Schwartz's theorem, we know that $u \in \mathbf{H}_k$ for some k. So $f + \lambda u \in \mathbf{H}_{\min(k,s)}$ for any λ. Choosing λ large enough, we conclude that $u = (L+\lambda I)^{-1}(f + \lambda u) \in \mathbf{H}_{\min(k+m,s+m)}$. If $k + m < s + m$ we can repeat

the argument to conclude that $u \in \mathbf{H}_{\min(k+2m,s+m)}$. we can keep going until we conclude that $u \in \mathbf{H}_{s+m}$. QED

Notice as an immediate corollary that any solution of the homogeneous equation $Lu = 0$ is C^∞.

We now obtain a second important consequence of Proposition 2.8.2. Choose λ so large that the operators

$$(L + \lambda I)^{-1} \quad \text{and} \quad (L^* + \lambda I)^{-1}$$

exist as operators from $\mathbf{H}_0 \to \mathbf{H}_m$. Follow these operators with the injection $\iota_m : \mathbf{H}_m \to \mathbf{H}_0$ and set

$$M := \iota_m \circ (L + \lambda I)^{-1}, \quad M^* := \iota_m \circ (L^* + \lambda I)^{-1}.$$

Since ι_m is compact (Rellich's lemma) and the composite of a compact operator with a bounded operator is compact, we conclude

Theorem 2.8.2 *The operators M and M^* are compact.*

Suppose that $L = L^*$. (This is usually expressed by saying that L is "formally self-adjoint". More on this terminology will come later.) This implies that $M = M^*$. In other words, M is a compact self adjoint operator, and we can apply Theorem 2.3.1 to conclude that eigenvectors of M form a basis of $R(M)$ and that the corresponding eigenvalues tend to zero. Prop 2.8.2 says that $R(M)$ is the same as $\iota_m(\mathbf{H}_m)$ which is dense in $\mathbf{H}_0 = L_2(\mathbf{T})$. We conclude that the eigenvectors of M form a basis of $L_2(\mathbf{T})$. If $Mu = ru$ then $u = (L + \lambda I)Mu = rLu + \lambda ru$ so u is an eigenvector of L with eigenvalue

$$\frac{1 - r\lambda}{r}.$$

We conclude that the eigenvectors of L are a basis of $\mathbf{H}_0$. We claim that only finitely many of these eigenvalues of L can be negative. Indeed, since we know that the eigenvalues r_n of M approach zero, the numerator in the above expression is positive, for large enough n, and hence if there were infinitely many negative eigenvalues μ_k, they would have to correspond to negative r_k and so these $\mu_k \to -\infty$. But taking $s_k = -\mu_k$ as the λ in (2.38) in Prop. 2.8.1 we conclude that $u = 0$, if $Lu = \mu_k u$ if k is large enough, contradicting the definition of an eigenvector. So all but a finite number of the r_n are positive, and these tend to zero. To summarize:

Theorem 2.8.3 *The eigenvectors of L are C^∞ functions which form a basis of $\mathbf{H}_0$. Only finitely many of the eigenvalues μ_k of L are negative and $\mu_n \to \infty$ as $n \to \infty$.*

It is easy to extend the results obtained above for the torus in two directions. One is to consider functions defined in a **domain** = bounded open set $\mathcal{G}$ of $\mathbf{R}^n$ and the other is to consider functions defined on a compact manifold. In both cases a few elementary tricks allow us to reduce to the torus case. We sketch what is involved for the manifold case.

2.9 Extension of the basic lemmas to manifolds.

Let $E \to M$ be a vector bundle over a manifold. We assume that M is equipped with a density which we shall denote by $|dx|$ and that E is equipped with a positive definite (smoothly varying) scalar product, so that we can define the L_2 norm of a smooth section s of E of compact support:

$$\|s\|_0^2 := \int_M |s|^2(x)|dx|.$$

Suppose for the rest of this section that M is compact. Let $\{U_i\}$ be a finite cover of M by coordinate neighborhoods over which E has a given trivialization, and ρ_i a partition of unity subordinate to this cover. Let ϕ_i be a diffeomorphism or U_i with an open subset of $\mathbf{T}^n$ where n is the dimension of M. Then if s is a smooth section of E, we can think of $(\rho_i s)\circ\phi_i^{-1}$ as an $\mathbf{R}^m$ or $\mathbf{C}^m$ valued function on $\mathbf{T}^n$, and consider the sum of the $\|\cdot\|_k$ norms applied to each component. We shall continue to denote this sum by $\|\rho_i f \circ \phi_i^{-1}\|_k$ and then define

$$\|f\|_k := \sum_i \|\rho_i f \circ \phi_i^{-1}\|_k$$

where the norms on the right are in the norms on the torus. These norms depend on the trivializations and on the partitions of unity. But any two norms are equivalent, and the $\| \|_0$ norm is equivalent to the "intrinsic" L_2 norm defined above. We define the Sobolev spaces $\mathbf{W}_k$ to be the completion of the space of smooth sections of E relative to the norm $\| \|_k$ for $k \geq 0$, and these spaces are well defined as topological vector spaces independently of the choices. Since Sobolev's lemma holds locally, it goes through unchanged. Similarly Rellich's lemma: if s_n is a sequence of elements of $\mathbf{W}_\ell$ which is bounded in the $\| \|_\ell$ norm for $\ell > k$, then each of the elements $\rho_i s_n \circ \phi_i^{-1}$ belong to $\mathbf{H}_\ell$ on the torus, and are bounded in the $\| \|_\ell$ norm, hence we can select a subsequence of $\rho_1 s_n \circ \phi_1^{-1}$ which converges in $\mathbf{H}_k$, then a subsubsequence such that $\rho_i s_n \circ \phi_i^{-1}$ for $i = 1, 2$ converge etc. arriving at a subsequence of s_n which converges in $\mathbf{W}_k$.

A differential operator L mapping sections of E into sections of E is an operator whose local expression (in terms of a trivialization and a coordinate chart) has the form

$$Ls = \sum_{|p|\leq m} \alpha_p(x) D^p s$$

Here the a_p are linear maps (or matrices if our trivializations are in terms of $\mathbf{R}^m$).

Under changes of coordinates and trivializations the change in the coefficients are rather complicated, but the **symbol** of the differential operator

$$\sigma(L)(\xi) := \sum_{|p|=m} a_p(x)\xi^p \qquad \xi \in T^*M_x$$

is well defined.

If we put a Riemann metric on the manifold, we can talk about the length $|\xi|$ of any cotangent vector.

If L is a differential operator from E to itself (i.e. F=E) we shall call L **even elliptic** if m is even and there exists some constant C such that

$$\langle v, \sigma(L)(\xi)v\rangle \geq C|\xi|^m|v|^2$$

for all $x \in M,\ v \in E_x,\ \xi \in T^*M_x$ and $\langle\ ,\ \rangle$ denotes the scalar product on E_x. Gårding's inequality holds. Indeed, locally, this is just a restatement of the (vector valued version) of Gårding's inequality that we have already proved for the torus. But Stage 4 in the proof extends unchanged (other than the replacement of scalar valued functions by vector valued functions) to the more general case.

2.10 Example: Hodge theory.

We assume knowledge of the basic facts about differentiable manifolds, in particular the existence of an operator $d : \Omega^k \to \Omega^{k+1}$ with its usual properties, where Ω^k denotes the space of exterior k-forms. Also, if M is orientable and carries a Riemann metric then the Riemann metric induces a scalar product on the exterior powers of T^*M and also picks out a volume form. So there is an induced scalar product $(\ ,\ ) = (\ ,\ )_k$ on Ω^k and a formal adjoint δ of d

$$\delta :\ \Omega^k \to \Omega^{k-1}$$

and satisfies

$$(d\psi, \phi) = (\phi, \delta\phi)$$

where ϕ is a $(k+1)$-form and ψ is a k-form. Then

$$\Delta := d\delta + \delta d$$

is a second order differential operator on Ω^k and satisfies

$$(\Delta\phi, \phi) = \|d\phi\|^2 + \|\delta\phi\|^2$$

where $\|\phi\|\|^2 = (\phi, \phi)$ is the intrinsic L_2 norm (so $\|\ \| = \|\ \|_0$ in terms of the notation of the preceding section). Furthermore, if

$$\phi = \sum_I \phi_I dx^I$$

is a local expression for the differential form ϕ, where

$$dx^I = dx_{i_1} \wedge \cdots \wedge dx_{i_k} \quad I = (i_1, \ldots, i_k)$$

then a local expression for Δ is

$$\Delta\phi = -\sum g^{ij} \frac{\partial \phi_I}{\partial x^i \partial x^j} + \cdots$$

where

$$g^{ij} = \langle dx^i, dx^j \rangle$$

and the $\cdots$ are lower order derivatives. In particular Δ is elliptic.

Let $\phi \in \Omega^k$ and suppose that

$$d\phi = 0.$$

Let $\mathcal{C}(\phi)$, the **cohomology class** of ϕ be the set of all $\psi \in \Omega^k$ which satisfy

$$\phi - \psi = d\alpha, \quad \alpha \in \Omega^{k-1}$$

and let

$$\overline{\mathcal{C}(\phi)}$$

denote the closure of $\mathcal{C}$ in the L_2 norm. It is a closed subspace of the Hilbert space obtained by completing Ω^k relative to its L_2 norm. Let us denote this space by L_2^k, so $\overline{\mathcal{C}(\phi)}$ is a closed subspace of L_2^k.

Proposition 2.10.1 *If $\phi \in \Omega^k$ and $d\phi = 0$, there exists a unique $\tau \in \overline{\mathcal{C}(\phi)}$ such that*

$$\|\tau\| \leq \|\psi\| \quad \forall\ \psi \in \mathcal{C}(\phi).$$

Furthermore, τ is smooth,and

$$d\tau = 0 \quad \textit{and} \quad \delta\tau = 0.$$

If choose a minimizing sequence for $\|\psi\|$ in $\mathcal{C}(\phi)$.

If we choose a minimizing sequence for $\|\psi\|$ in $\mathcal{C}(\phi)$ we know it is Cauchy, cf. the proof of the existence of orthogonal projections in a Hilbert space. So we know that τ exists and is unique. For any $\alpha \in \Omega^{k+1}$ we have

$$(\tau, \delta\alpha) = \lim(\psi, \delta\alpha) = \lim(d\psi, \alpha) = 0$$

as ψ ranges over a minimizing sequence. The equation $(\tau, \delta\alpha) = 0$ for all $\alpha \in \Omega^{k+1}$ says that τ is a weak solution of the equation $d\tau = 0$.

We claim that

$$(\tau, d\beta) = 0 \quad \forall\ \beta \in \Omega^{k-1}$$

which says that τ is a weak solution of $\delta\tau = 0$. Indeed, for any $t \in \mathbf{R}$,

$$\|\tau\|^2 \leq \|\tau + td\beta\|^2 = \|\tau\|^2 + t^2\|d\beta\|^2 + 2t(\tau, d\beta)$$

so

$$-2t(\tau, d\beta) \leq t^2\|d\beta\|^2.$$

If $(\tau, d\beta) \neq 0$, we can choose

$$t = -\epsilon\frac{(\tau, d\beta)}{|(\tau, d\beta)|}, \quad \epsilon > 0$$

so

$$|(\tau, d\beta)| \leq \epsilon |d\beta|^2.$$

As ϵ is arbitrary, this implies that $(\tau, d\beta) = 0$.

So $(\tau, \Delta\psi) = (\tau, [d\delta + \delta d]\psi) = 0$ for any $\psi \in \Omega^k$. Hence τ is a weak solution of $\Delta\tau = 0$ and so is smooth. The space $\mathcal{H}^k$ of weak, and hence smooth solutions of $\Delta\tau = 0$ is finite dimensional by the general theory. It is called the space of Harmonic forms. We have seen that there is a unique harmonic form in the cohomology class of any closed form, s the cohomology groups are finite dimensional. In fact, the general theory tells us that

$$L_2^k \bigoplus_\lambda E_\lambda^k$$

(Hilbert space direct sum) where E_λ^k is the eigenspace with eigenvalue λ of Δ. Each E_λ is finite dimensional and consists of smooth forms, and the $\lambda \to \infty$. The eigenspace E_0^k is just $\mathcal{H}^k$, the space of harmonic forms. Also, since

$$(\Delta\phi, \phi) = \|d\phi\|^2 + \|\delta\phi\|^2$$

we know that all the eigenvalues λ are non-negative.

Since $d\Delta = d(d\delta + \delta d) = d\delta d = \Delta d$, we see that

$$d : E_\lambda^k \to E_\lambda^{k+1}$$

and similarly

$$\delta : E_\lambda^k \to E_\lambda^{k-1}.$$

For $\lambda \neq 0$, if $\phi \in E_\lambda^k$ and $d\phi = 0$, then $\lambda\phi = \Delta\phi = d\delta\phi$ so $\phi = d(1/\lambda)\delta\phi$ so d restricted to the E_λ is exact, and similarly so is δ. Furthermore, on $\bigoplus_k E_\lambda^k$ we have

$$\lambda I = \Delta = (d + \delta)^2$$

so we have

$$E_\lambda^k = dE_\lambda^{k-1} \oplus \delta E_\lambda^{k+1}$$

and this decomposition is orthogonal since $(d\alpha, \delta\beta) = (d^2\alpha, \beta) = 0$.

As a first consequence we see that

$$L_2^k = \mathcal{H}^k \oplus \overline{d\Omega^{k-1}} \oplus \overline{\delta\Omega^{k-1}}$$

(the Hodge decomposition). If H denotes projection onto the first component, then Δ is invertible on the image of $I - H$ with a an inverse there which is compact. So if we let N denote this inverse on im $I - H$ and set $N = 0$ on $\mathcal{H}^k$ we get

$$\begin{aligned} \Delta N &= I - H \\ Nd &= dN \\ \delta N &= N\delta \\ \Delta N &= N\Delta \\ NH &= 0 \end{aligned}$$

which are the fundamental assertions of Hodge theory, together with the assertion proved above that $H\phi$ is the unique minimizing element in its cohomology class.

We have seen that

$$d+\delta:\ \bigoplus_k E^{2k}_\lambda \to \bigoplus_k E^{2k+1}_\lambda \ \text{ is an isomorphism for } \lambda \neq 0 \tag{2.39}$$

which of course implies that

$$\sum_k (-1)^k \dim E^k_\lambda = 0$$

This shows that the index of the operator $d+\delta$ acting on $\bigoplus L^k_2$ is the Euler characteristic of the manifold. (The index of any operator is the difference between the dimensions of the kernel and cokernel).

Let $P_{k,\lambda}$ denote the projection of L^k_2 onto E^k_λ. So

$$e^{-t\Delta} = \sum e^{-\lambda t} P_{k,\lambda}$$

is the solution of the heat equation on L^k_2. As $t \to \infty$ this approaches the operator H projecting L^k_2 onto $\mathcal{H}_k$. Letting Δ_k denote the operator Δ on L^k_2 we see that

$$\operatorname{tr} e^{-t\Delta_k} = \sum e^{-\lambda_k}$$

where the sum is over all eigenvalues λ_k of Δ_k counted with multiplicity. It follows from (2.39) that the alternating sum over k of the corresponding sum over non-zero eigenvalues vanishes. Hence

$$\sum (-1)^k \operatorname{tr} e^{-t\Delta_k} = \chi(M)$$

is independent of t. The index theorem computes this trace for small values of t in terms of local geometric invariants.

The operator $d+\delta$ is an example of a Dirac operator whose general definition we will not give here. The corresponding assertion and local evaluation is the content of the celebrated Atiyah-Singer index theorem, one of the most important theorems discovered in the twentieth century.

2.11 The resolvent.

In order to connect what we have done here notation that will come later, it is convenient to let $A = -L$ so that now the operator

$$(zI - A)^{-1}$$

is compact as an operator on $\mathbf{H}_0$ for z sufficiently negative. (I have dropped the ι_m which should come in front of this expression.) The operator A now has only

finitely many positive eigenvalues, with the corresponding spaces of eigenvectors being finite dimensional. In fact, the eigenvectors $\lambda_n = \lambda_n(A)$ (counted with multiplicity) approach $-\infty$ as $n \to \infty$ and the operator $(zI - A)^{-1}$ exists and is a bounded (in fact compact) operator so long as $z \neq \lambda_n$ for any n. Indeed, we can write any $u \in \mathbf{H}_0$ as

$$u = \sum_n a_n \phi_n$$

where ϕ_n is an eigenvector of A with eigenvalue λ_n and the ϕ form an orthonormal basis of $\mathbf{H}_0$. Then

$$(zI - A)^{-1} u = \sum \frac{1}{z - \lambda_n} a_n \phi_n.$$

The operator $(zI - A)^{-1}$ is called the **resolvent** of A at the point z and denoted by

$$R(z, A)$$

or simply by $R(z)$ if A is fixed. So

$$R(z, A) := (zI - A)^{-1}$$

for those values of $z \in \mathbf{C}$ for which the right hand side is defined.

If z and are complex numbers with $\mathrm{Re} z > \mathrm{Re} a$, then the integral

$$\int_0^\infty e^{-zt} e^{at} dt$$

converges, and we can evaluate it as

$$\frac{1}{z - a} = \int_0^\infty e^{-zt} e^{at} dt.$$

If $\mathrm{Re} z$ is greater than the largest of the eigenvalues of A we can write

$$R(z, A) = \int_0^\infty e^{-zt} e^{tA} dt$$

where we may interpret this equation as a shorthand for doing the integral for the coefficient of each eigenvector, as above, or as an actual operator valued integral. We will spend a lot of time later on in this course generalizing this formula and deriving many consequences from it.